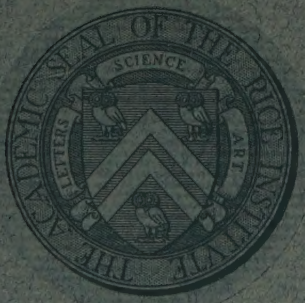
THE RICE INSTITUTE PAMPHLET

Vol. XX January, 1933 No. 1



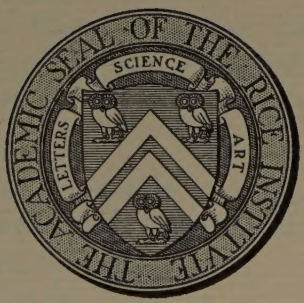
Published by

THE RICE INSTITUTE

A university of liberal and technical learning founded by William Marsh Rice in the City of Houston, Texas, and dedicated by him to the advancement of Letters, Science, and Art

THE
RICE INSTITUTE
PAMPHLET

Vol. XX January, 1933 No. 1



Published by
THE RICE INSTITUTE

A university of liberal and technical learning founded by William Marsh Rice in the City of Houston, Texas, and dedicated by him to the advancement of Letters, Science, and Art

THE RICE INSTITUTE PAMPHLET

VOL. XX January, 1933 No. 1

CONTENTS

PAGE

THE LATER SCIENTIFIC WORK OF HENRI POINCARÉ—Lectures delivered at the Rice Institute in May, 1925, with subsequent additions, by PROFESSOR J. HADAMARD, of the Collège de France and the École Polytechnique, in continuation of a series of lectures on the Early Scientific Work of Henri Poincaré, delivered at the Rice Institute in March, 1920, by Professor Hadamard, and published in the Rice Institute Pamphlet, Vol. IX, No. 3, July, 1922.

I. Introduction 1
II. The Theory of Centers 9
III. Equations of Higher Degree and the Intervention of Analysis Situs 29
IV. The Case of the Torus 42
V. Higher Differential Equations: Singular Points and the Theorem of Kronecker 55
VI. The Introduction of Periodic Solutions . . 59
VII. Work on Point Transformations after Poincaré 80

THREE PUBLIC LECTURES ON MODERN PHYSICS—Delivered at the Rice Institute in March, 1932, by PROFESSOR ARTHUR H. COMPTON of the University of Chicago.

I. What Things Are Made of 87
II. What Is Light? 102
III. The Riddle of Cosmic Rays 118

THE LATER SCIENTIFIC WORK OF HENRI POINCARÉ[1]

I

INTRODUCTION

HAPPY as I am to find myself again amongst you, and in the midst of the Rice Institute, the improvements and continuously increasing importance of which I follow with so much interest and pleasure, I am especially so to resume the subject to which our first colloquium of 1920 was devoted. Indeed, I have the feeling, a particularly agreeable one to me, that I am continuing today the last lecture which I delivered to you, on the occasion of my earlier visit in 1920. I must add that this idea was suggested to me by your President, to whom I am most grateful for his friendly insistence on it.

Moreover, no other subject could better deserve our keen and everlasting attention. There is hardly any limit to the interest we can take in reviewing and commenting on Poincaré's work and to the fruitful points of view we can find in it. Poincaré's work has not ceased to be the center of almost all our contemporary mathematical labor and perhaps centuries will pass before its full significance will be realized.

[1]Lectures delivered at the Rice Institute in May, 1925, by Professor J. Hadamard, of the Collège de France and the École Polytechnique, in continuation of a series of lectures on the Early Scientific Work of Henri Poincaré, delivered at the Rice Institute in March, 1920, by Professor Hadamard, and published in the Rice Institute Pamphlet, Vol. IX, No. 3, July, 1922.

A first part of that work, corresponding to those extraordinary first four or five years, the wonderful results of which have been the amazement of the scientific world, was examined before you in 1920.

To recall in a few words what we said in 1920, let me remind you that, as soon as we leave the well known elementary types, the treatment of differential equations can be, and has to be, pursued in several different directions. We can inquire whether the solution happens to admit of an expression by known functions, or try to express it by other classes of functions, by series or by other means taken from the classic methods of Calculus. Then, we can improve the local study, which is the fundamental discovery of Newton and Cauchy; and, since Cauchy, we can consider the required function, if analytic, from the point of view of the general theory of analytic functions.

We saw, in 1920, what Poincaré did in all these directions and how many corresponding important later works have been derived from his inspiration. However, I shall begin by reverting to two points of my previous exposé. In the first place, concerning integration by algebraic functions, I think it useful to mention the researches of Drach. This geometer, one of the most penetrating scientific minds of our time, has undertaken to find the algebraic solutions of a given algebraic differential equation of the first order and first degree

$$\text{(E)} \qquad \frac{dx}{X} = \frac{dy}{Y},$$

or, what amounts to the same, of the partial differential equation

$$\text{(Ea)} \qquad X\frac{\partial F}{\partial x} + Y\frac{\partial F}{\partial y} = 0,$$

where X, Y are given polynomials in x and y; but he remarks that a powerful help is given to such an investigation by a fact already used by Darboux, viz., that, if (E) is verified by the algebraic curve $f(x,y) = 0$, the function f, if irreducible, (which can always be assumed), must verify an identity of the form

$$X\frac{\partial f}{\partial x} + Y\frac{\partial f}{\partial y} = Mf,$$

where M is again a polynomial. The investigation of the latter polynomial, the degree of which is limited *a priori* (being less by one than the higher of the degrees of X and Y), is perhaps the easiest way to attack the problem, and enables Drach to improve considerably the results obtained by Poincaré in that line.

My second observation will concern Poincaré's paper in the *Bulletin Astronomique*, Vol. I, 319, devoted to the most general trigonometric series of the form

$$\sum_n (A_n \cos \alpha_n t + B_n \sin \alpha_n t). \tag{2}$$

We have previously mentioned, among the numerous and important researches inspired by Poincaré's work, the investigations of Bohl and Esclangon on quasi-periodic functions, which are derived from the memoir just quoted. I wish to say something more on this subject, on account of the fact that it has undergone an unexpected and remarkable development in recent times.

Bohl and Esclangon, starting from astronomical points of view such as those suggested by the memoir of Poincaré, are led to inquire what becomes of a function F of p variables

$$u_1, u_2, \cdots, u_p,$$

periodic in each of them, with periods $\omega_1, \omega_2, \cdots, \omega_p$, if we take the u's all equal to the same variable quantity t. If $p=1$, it is clear that we get again a periodic function of t, the period of which is ω_1. Moreover, a similar result would appear for $p=2$ if the two periods were commensurable. In the general case, the first question is whether $\omega_1, \cdots, \omega_p$ are "linearly dependent" or not, i.e., whether a linear homogeneous relation of the form

$$\frac{l_1}{\omega_1} + \cdots + \frac{l_p}{\omega_p} = 0, \tag{1}$$

with integral coefficients $l_1, \cdots, l_p$ (not all zero) can take place or not. In the first case, we are easily led to a similar problem with a smaller value of p; in the second, a remarkable importance attaches to a theorem of Kronecker,[1] according to which (under the hypothesis of linear independence) the inequalities

$$\left|\frac{\lambda_i}{\omega_i} - \mu_i - k_i\right| < \epsilon_i \qquad (i=1, 2, \cdots, p)$$

admit of common solutions in integers λ, μ, whatever $k_1, \cdots, k_p$ may be and however small we take the right-hand members. It results therefrom that the function F, by varying the single quantity t, practically assumes the same systems of values as if there were as many independent variables; and, in the second place, that such a function will admit of an infinity of approximate periods (whence its name). The properties of such functions, when submitted to quadratures or when appearing as coefficients in differential equations, are essential to Celestial Mechanics and have been profoundly investigated in the first place by the two above-named authors, but in more recent times and from

[1] A simple and elegant proof by Lettenmeyer has appeared in the *Proc. London Math. Soc.*, XXI (1923), 306.

a more general point of view, by quite another group of geometers.

This is, indeed, one of the most remarkable instances of the profound and rather mysterious unity of science. While quasi-periodic functions were introduced, as we have just seen, by the necessities of Analytical and Celestial Mechanics, "almost periodic" functions, their close allies and natural generalization, were suggested, with the same degree of necessity, by the needs of Higher Arithmetic. I refer to the theory of Prime Numbers, as we regard it since Dirichlet and Riemann, viz., in connection with the properties of Dirichlet's series. In the hands of Harald Bohr and his disciples, these almost periodic functions have considerably improved our knowledge of Riemann's and allied transcendentals.

An almost periodic function $f(t)$ is, at least formally, represented, and in every case, defined by a Fourier-like expansion

$$\Sigma A_n e^{i\alpha_n t}, \tag{2}$$

where the exponents α_n are not subjected (as in ordinary Fourier's series) to being multiples of one and the same quantity, nor—as in quasi-periodic functions—to be derived from a finite basis $\left(\frac{2\pi}{\omega_1}, \cdots, \frac{2\pi}{\omega_p}\right)$, i.e., to be linear combinations of $\frac{2\pi}{\omega_1}, \cdots, \frac{2\pi}{\omega_p}$, with integral coefficients. The first fundamental result of this theory has been to prove that such functions f are again fully characterized by the existence of an infinite number of "approximate periods" or "translation numbers" τ connected with the approximation ϵ in such a way that we have for every t

$$|f(t+\tau) - f(t)| < \epsilon, \tag{3}$$

with the supplementary condition that, for any given ϵ, there exists a length T such that every t-interval of magnitude T contains such an approximate period satisfying the above inequality for every t. This set of conditions (the latter of which is essential) is sufficient in order that we have an almost periodic function in the above meaning; the proof of this is, however, notably more difficult than in the quasi-periodic case, where the approximate periods were known in advance on account of their connection with the given basis. Bohr's proof rests on a profound generalization of Parseval's known theorem concerning the integral of the square of a trigonometric series, which generalization affords the sum of certain infinite numbers of positive terms: it is not even known in advance whether this infinity is a numerable one, but this precisely results from the fact that the sum is finite, and the different terms correspond to those of the series (2).

Seeing the importance, in recent years, of this theory of Bohr and of its arithmetical applications, one might believe that, with its introduction, we had deserted the domain of Dynamics for the domain of the Theory of Numbers. But this would be to doubt, quite unjustly, the unity of science; in fact, the connection of Bohr's theory with Analytical Mechanics and with the development of Poincaré's ideas in that field, has now been made quite evident; so that, in this way, Arithmetic and Dynamics are becoming most closely united.

We said in 1920 that besides such points of view as we have just recalled, i.e., expressions of solutions by known classes of functions, Poincaré had adopted quite different ones, hardly thought of before him, consisting of the separate and individual study of the real domain, a study in which the coefficients of the differential equation are not

even assumed always to be analytic, and, as we noticed, an exceptionally difficult one on account of the fact that we are deprived of the powerful help afforded by the theory of analytic functions. Let us recall also[1] that the profound and intimate principle of his methods may be considered as being inspired by the example of modern Algebra, and consists in introducing simultaneously and comparing with each other several solutions of the equation, instead of considering one of them individually. We shall see that the same observation will hold in what we are going to say this time.

It must be noticed that this new view of the question requires a completion of the classic existence theorem for differential equations, the mode of dependence of the solution, not only with respect to the main variable, but also with respect to the initial data, requiring investigation. This has been done, of course, by Poincaré, in his *Méthodes nouvelles de la Mécanique Céleste*, at least for the analytic case, the dependence being itself an analytic one; the general case gives an analogous result, as can be easily shown.[2]

The first and second parts of the *Mémoire sur les courbes définies par les équations différentialles*[3] have been considered by us as belonging to Poincaré's earlier work and analyzed in 1920. The third and fourth ones[4] will be considered

[1]See my previous Lectures of 1920, p. 155.

The method of considering the solution of a differential equation as a function not only of the independent variable, but also of the initial conditions, is now becoming quite common in contemporary science, since Poincaré thought of it. For instance, researches derived indirectly, but certainly, from the inspiration of Poincaré are those of K. Popoff, who has tried to expand the solution of the ballistic problem not in terms of the time, but in terms of the angular variable which gives the initial direction, and has dealt similarly with the equations of Celestial Mechanics.

[2]See, for instance, my *Cours d'Analyse*, vol. II, No. 243*bis*, 315.

[3]*Jour. de Math.*, ser. 3, 7 (1881), 375–422; ser. 3, 8 (1882), 251–296.

[4]*Ibid.*, ser. 4, 1 (1885), 167–244; ser. 4, 2 (1886), 151–217.

now. It may seem odd that in separating his earlier from his later work, we make the division between two parts of the same Memoir or series of Memoirs. The reason, which will appear, I hope, altogether clear to you, is that, in the latter parts, Poincaré uses quite new means of action, methods which were not employed in the earlier parts, and which constitute quite a new moment in mathematical science.

II

THE THEORY OF CENTERS

WE return to the equation of the first order and first degree

$$(1) \qquad \frac{dx}{X(x,y)} = \frac{dy}{Y(x,y)} \quad \text{or} \quad (1') \ \frac{dx}{dt} = X, \frac{dy}{dt} = Y.$$

We have seen[1] that the problem is to be formulated in the following geometrical way on a plane or, as Poincaré often does,[2] on a sphere: with each point $M(x, y)$ on the surface (i.e., the point with cartesian coordinates x, y on the plane or a point of the sphere corresponding to it by central projection), we associate a vector issuing from M (and lying in the plane, or in the tangent plane to the sphere at M), and whose direction will be given by (1). Then, a solution of the problem is a curve of the surface whose tangent at this point is the vector corresponding to that point.

In this definition of the vector field, the direction and not the magnitude of the vector is important. Any vector with components proportional to X and Y can replace the original one. When however, $X = Y = 0$, there is no direction defined, the vector being zero. Around such a point O, the direction of the vector is likely not to be continuous, and O is a *singular point* of the differential equation. Following Poincaré and treating, with him, the most obvious

[1] See my previous Lectures, p. 172.

[2] *Loc. cit.*, taking account of the corresponding footnote.

and general case, in which the expansions of X,Y in powers of x,y contain linear terms, we have seen[1] that he found four usual classes of singular points, differentiated according to the behavior of the solutions in their neighborhoods. They are sufficiently explained by the following four figures. These various cases, however, do not occur

1. Node 2. Pass 3. Focus 4. Center

with the same generality. Case 4, together with 2, is the usual case when the equation is integrable, but it is highly exceptional in the general case. The general cases are 1, 2, and 3.

The distinction between the above cases depends on two numbers s and s', the roots of a certain quadratic equation which can be deduced from the linear terms of X and Y. The node arises when the roots are real and of the same sign; the pass, when the roots are real and of opposite signs; the focus, when s and s' are complex (being conjugate to each other).

When s and s' are pure (and opposite) imaginaries, and only then, the singular point *may be* a center. This already shows that the case of a center is an exceptional one. But, in fact,—and this is what Poincaré now investigates—we are going to see that this condition of s and s' being pure imaginaries, which is necessary for having a center, is far from being sufficient.

[1] *Loc. cit.*, p. 174.

The question can be faced in the following way. Let us take the starting point m of our trajectory on a determinate direction issuing from the singular point O: for instance, the direction of positive x's. In the present case, the curve will turn around O and intersect again the positive part of the x-axis at m', which we could call a "consequent" of m. If this new point m' coincides with m, the trajectory will be a closed one and if this happens for every position of m sufficiently near to O, O will be a center.[1] If not,[2] we shall have a diagram such as that represented in fig. 1 (p. 176) of our previous Lectures, and O will be a focus,[3] the trajectory either approaching O asymptotically as t increases or receding more and more from it.

This is what Poincaré is now going to discuss, under the hypothesis that X and Y are holomorphic around O:—a hypothesis which, in this case, seems to be an essential one, it being doubtful whether a satisfactory solution can ever be found in the non-analytic case. The method will consist in trying to find, for the given equation, an integral

$$F(x, y) = C,$$

F being required to satisfy the partial differential equation

$$X\frac{\partial F}{\partial x} + Y\frac{\partial F}{\partial y} = 0. \tag{2}$$

[1]In the non-analytic case, Bendixson has proposed a slightly different definition; see below, p. 24.

[2]The segment mm' is a function of the length Om, which becomes infinitesimal with Om. It is an infinitesimal of the first order in the general case of a focus, when s and s' are complex; of higher order in the case which we are now dealing with.

[3]If m is not sufficiently near to O, and m' lies inside m, the trajectory, instead of tending to O, may admit of a limiting cycle (see my previous Lectures, p. 179–180); but, *in the analytic case* (Cf. p. 24), this cannot be when Om is sufficiently small.

In the present case, it may be assumed that X, Y respectively begin with the linear terms $y, -x$

$$X = y + X_2 + X_3 + \cdots, \quad Y = -x + Y_2 + Y_3 + \cdots,$$

(where X_q, Y_q are homogeneous polynomials of degree q in x,y); then the above partial differential equation can be written

$$(3) \quad y\frac{\partial F}{\partial x} - x\frac{\partial F}{\partial y} = -(X_2 + X_3 + \cdots)\frac{\partial F}{\partial x} - (Y_2 + Y_3 + \cdots)\frac{\partial F}{\partial y}.$$

If F, assumed to be holomorphic is also written in the form

$$F = F_2 + F_3 + \cdots,$$

F_q again denoting the homogeneous part of order q of the expansion and F_2 being immediately found to be $x^2 + y^2$, we see that F_q will appear in terms of order q in the left-hand member of (3) and only in higher terms in the right-hand member, so that the F's can be determined in the usual way, by successive equations of the form

$$(3') \quad y\frac{\partial F_q}{\partial x} - x\frac{\partial F_q}{\partial y} = H_q,$$

where each H_q (again a homogeneous polynomial of an order denoted by its suffix) only contains those polynomials F the suffixes of which are less than q, which may be assumed to have been previously calculated. Introducing polar coordinates by $x = \rho \cos \omega$, $y = \rho \sin \omega$, we write F in the form

$$(4) \quad F = \rho^2 + F_3\rho^3 + F_4\rho^4 + \cdots,$$

where F_q, for the sake of simplicity, is written instead of $F_q(\cos \omega, \sin \omega)$ and depends upon the sines and cosines

of $q\omega$, $(q-2)\omega$, $(q-4)\omega$, $\cdots$. We have, on making the substitution, the successive equations

$$\frac{dF_3}{d\omega} = H_3 = H_3(\cos\omega, \sin\omega), \tag{5}$$

$$\frac{dF_4}{d\omega} = H_4 = H_4(\cos\omega, \sin\omega), \cdots,$$

where H_q has the properties mentioned above for F_q. These are easy to integrate. Thus

$$\frac{dF_3}{d\omega} = H_3 = -2(X_2\cos\omega + Y_2\sin\omega)$$
$$= A\cos\omega + B\sin\omega + C\cos 3\omega + D\sin 3\omega$$

gives

$$F_3 = A\sin\omega - B\cos\omega + \frac{C}{3}\sin 3\omega - \frac{D}{3}\cos 3\omega,$$

where no constant term is to be added, as only odd multiples of ω must appear in F_3.

Then, we are able to construct H_4, viz.,

$$H_4 = -2(X_3\cos\omega + Y_3\sin\omega) + (3F_3\cos\omega + \frac{dF_3}{d\omega}\sin\omega)X_2$$
$$+ (3F_3\sin\omega - \frac{dF_3}{d\omega}\cos\omega)Y_2,$$

and we immediately see that this expression, when expressed as a trigonometric polynomial in ω, satisfies the required condition concerning parisymmetry, i.e., contains only terms corresponding to even multiples of ω (not greater than 4):

$$H_4 = C_0 + C_2\cos 2\omega + D_2\sin 2\omega + C_4\cos 4\omega + D_4\sin 4\omega. \tag{5'}$$

But here, two cases are possible. In general, C_0 will not be zero. If so, the quantity F_4, deduced from H_4 by integration with respect to ω, cannot be purely trigonometric: it will contain a term proportional to ω itself. We can call such a term a *secular* one: for such a calculation is obviously quite analogous to what is classic in Celestial

Mechanics, where the presence of these secular terms is a well known difficulty.

If such is the case here, we see that we cannot find for F an expression of the required form. Are we, then, allowed to say that the origin is certainly *not* a center? At first, this seems to be doubtful. Indeed, a similar conclusion would not be legitimate in the classic case of Celestial Mechanics: it would be quite possible, there, that a term of the form $C\omega$, associated with other ones in $\omega^3, \omega^5, \cdots$ (such as would appear by pursuing the calculation) could be, for instance, the expansion of

$$\sin C\omega = C\omega - \frac{C^3\omega^3}{6} + \cdots.$$

To prove rigorously that matters stand differently in the present problem, and that the case of a center is actually excluded when $C_0 \neq O$, is therefore not an insignificant difficulty. But we are able to master it thanks to the principles already introduced by Poincaré in the earlier parts of the Memoir. Indeed, under the above assumption, we can construct, around the origin, *a family of contactless cycles*.[1] We do this by substituting for (5) the new equation

$$(6) \qquad \frac{dG}{d\omega} = H_4 - C_0 = C_2\cos 2\omega + D_2\sin 2\omega + C_4\cos 4\omega + D_4\sin 4\omega,$$

and noticing that this gives for G_4 a purely trigonometric function of ω. Now, the quantity

$$(7) \qquad G = \rho^2 + F_3\rho^3 + G_4\rho^4$$

is a function of x,y (viz., a polynomial in x,y) which, being substituted in the left hand side of (2) will give an expression of the form

[1]See my previous Lectures, p. 182.

$$(7') \quad C_0\rho^4 + \text{terms of higher order} = C_0(x^2+y^2)^2 + \text{terms of higher order},$$

the value of which will be constantly positive throughout a certain circle having O for its center, if, for instance, we suppose C_0 to be positive. G is what we could call an "unintegral," if the word would not look too barbarous: in fact, it has to integrals of our differential equation, the same relation as inequalities to equalities. We have already seen, in our previous Lectures how such "unintegrals" are of constant use in researches of the present kind. This is again the case here: the existence of such an "unintegral" evidently excludes the possibility of any closed trajectory in the neighborhood of the origin. Thus, the required conclusion is proved in full rigor.

It is easy to see that, under such circumstances, O must be a focus. For, as G is constantly increasing when t increases, and remains finite, it must have a limit when t tends to infinity. Moreover, dG/dt must tend to zero,[1] which can only be, if it is the case for ρ (while on the other hand, we know that ω is indefinitely increasing).

[1]This may appear as intuitive in the present case. But, generally speaking, we have the theorem, which has been recognized by Kneser as well as myself (1896) to be of great use in these studies, that *if a function G of t tends to a finite limit when $t=\infty$ and the second derivative $G''(t)$ remains finite, then the first derivative $G'(t)$ tends to zero.* The same theorem has been found again by Littlewood, *Proc. London Math. Soc.*, IX_2 (1910–11); also, in a more special case, by C. N. Moore, *Trans. Am. Math. Soc.*, VIII (1907), in connection with arithmetical researches, after which Hardy and Littlewood and several authors, especially Landau and recently Mordell (see also my own paper in the *Bull. Soc. Math. France, Proc. Verb. des séances*, 1914, 68) generalized it by showing that it corresponds to inequalities between the maxima of successive derivatives of an arbitrary function of a real variable.

Of course, the condition concerning G'' is fulfilled here, as

$$G''(t) = X\frac{\partial G'}{\partial x} + Y\frac{\partial G'}{\partial y},$$

G' being the expression (7').

Let us now suppose that C_0 is zero, so that F_4 can be constructed, and even in an infinity of ways, as we can add an arbitrary constant c, giving a term $c(x^2+y^2)$ in the value of F. We shall find no difficulty in the construction of F_5, which will be quite analogous to that of F_3; but an impossibility will again arise in the construction of F_6, if the constant term analogous to C_0 in the corresponding H_6 be different from zero: if so, we shall be able to construct an "unintegral" G, the derivative of which will have an expression analogous to (7′) but, beginning with a term in $(x^2+y^2)^3$, whence we can again conclude that the origin is necessarily a focus. We can go on in the same way, the calculation coming to an end if we meet with an H (necessarily of an even suffix) whose constant term—in other words, whose mean value with respect to ω in the interval $(0, 2\pi)$—is different from zero, in which case we can assert that O is a focus.

I shall, however, insist on one detail which Poincaré does not mention explicitly, so that it might appear to leave room for an objection to his argument, which is, in reality, perfectly valid and rigorous. We have noticed that, if F_4 can be found, its value contains an arbitrary constant c. Of course, the value of c, if left at first indeterminate, will occur in every further calculation; therefore, at a first glance, there seems to be no reason why the behaviour of the calculation of F_6 (or of $F_8, \cdots$) should not depend on that value of c, the constant term in H_6 (or H_8, etc.) being zero if and only if c is chosen in a proper way. A similar objection would seemingly arise from the presence of arbitrary constants in F_6, F_8, etc. That such will never be the case, is a consequence of the above discussion. The possibility of constructing F_6, for instance, shows, as is easy to see, that the segment intercepted on a radius vector between the

initial point of a trajectory and the point which is found after turning once around O,—i.e., the difference between the initial radius vector ρ_0 corresponding to $\omega=0$ and the radius vector ρ_1 corresponding to $\omega=2\pi$, is of the 7th order in ρ_0, while it would be of order 5 if the constant term of H_6 were different from zero: it is clear that this character only depends on the given question, and not on our ways of performing the calculations. But we can see the same otherwise, by noticing that, if the integral F actually exists, the quantity

$$\bar{F}=\chi(F)=F+cF^2+c_1F^3+\cdots \tag{8}$$

will admit of the same property, whatever the function χ or, consequently, whatever the coefficients c, c_1, $\cdots$ may be; and we can say the same as concerns the fact of F being an integral in the first, the second, $\cdots$ approximation, i.e., such that $\chi(F)$ begins with terms in ρ^4, ρ^6, $\cdots$. Now, the constant c in expression (8) is precisely the additive constant in F_4 which we have considered above and denoted by the same letter. A similar observation applying to higher coefficients in the right-hand side of (8), we see that choosing arbitrarily the additive constants in the successive F_{2n}'s comes to the same as varying arbitrarily such coefficients, and cannot change the possibility or impossibility of the successive calculations.

We can, for example, say that we take every F_{2n} without a constant term in its trigonometric expansion; or also, as Poincaré will be led to do, that we take it such that it vanishes for $\omega=0$.

Now, what will happen, if, indefinitely, the H_{2n}'s are devoid of constant terms, so that every F_{2n} can be constructed? A formal expression of F is thus obtained. The question is whether it will be convergent or, more exactly

(according to what we have seen above), whether we can choose the integration constants in the successive F_{2n}'s so as to make it convergent.

Before examining this, let us observe that the successive relations (3′), (5) could have been found by writing our differential equation in polar coordinates, in the form

$$\frac{d\rho}{d\omega} = -\Psi(\rho, \omega) = -(\rho^2\theta_2(\omega) + \rho^3\theta_3(\omega) + \cdots), \tag{9}$$

which gives for the partial differential equation (2)

$$\frac{\partial F}{\partial \omega} - \Psi \frac{\partial F}{\partial \rho} = 0, \tag{9a}$$

and that, conversely, every equation of the form (9a) could be treated in the same way as (1), with the only difference that the functions F_3, F_4, $\cdots$, which would appear as coefficients of the successive powers of ρ in F would no longer be periodic. The calculation would lead us to successive H's, whence the F_m's would have to be deduced let us say for $0 \leq \omega \leq 2\pi$, by integration with respect to ω: for instance, we could take 0 as the lower limit for each integration.

In particular what should be obtained by treating precisely in this way our original equation (1)? Such a treatment will be different from the above: for, up to now, odd terms in (4) have been taken so as to have zero for their mean value, and not so as to be zero for $\omega = 0$. In the new mode of calculation, the above enunciated conditions concerning parisymmetry will not be verified, so that multiples of ω of every parity may appear in any F. Of course, we may again fear that the calculation thus modified might bring in secular terms and, this time, in odd coefficients as well as in even ones; but the objection can be answered

in just the same way as before: from our integral expansion F (as formerly constructed), we can deduce the new one

$$F+\gamma_1 F^{3/2}+\gamma_2 F^2+\cdots$$

where the γ's are again arbitrary coefficients. In a manner similar to what we have seen above, the choice of γ_1 is equivalent to the introduction of an arbitrary constant in F_3; this being done in a determinate way, the choice of γ_2 will be equivalent to the introduction of an arbitrary constant in F_4; and so on, whence we again conclude that the appearance or non-appearance of secular terms is utterly independent of the values of such constants.

We, therefore, can choose these so that every F_q vanishes for $\omega=0$; and this will allow us to apply the method of dominating functions. For, if thus conducted, the calculation of the successive H's will introduce only additions, multiplications and integrations with the lower limit zero. Therefore, if we replace the expansion of Ψ by a dominating one, the new expressions of the H's will throughout dominate the original one.

Now, this expansion of Ψ in powers of ρ is assumed to be convergent, and to be so whatever ω may be (if only real) in an interval $(0, \alpha)$ of values of ρ. There is no loss of generality in admitting that α is greater than 1, as this can always be obtained by a homothetic transformation: on account of which, in (9) all the θ's will be, for every ω between 0 and 2π, smaller in absolute value than a positive fixed number K and the expansion (9) will be dominated[1] by

$$\frac{K\rho^2}{1-\rho}. \tag{10}$$

[1]Poincaré himself assumes X and Y to be polynomials: which leads him to dominate (9) by $\frac{K(\rho^2+\rho^3+\cdots+\rho^p)}{1-K(\rho+\rho^2+\cdots+\rho^q)}$ from which he gets to (10). But there is no difficulty whatever in taking for X, Y indefinite Maclaurin expansions.

The question is therefore whether the differential equation

$$\frac{d\rho}{d\omega} = -\frac{K\rho^2}{1-\rho} \tag{11}$$

admits an integral, or the corresponding partial differential equation

$$\frac{\partial F}{\partial \omega} = \frac{K\rho^2}{1-\rho}\frac{\partial F}{\partial \rho}, \tag{11a}$$

a solution which reduces to ρ^2 for $\omega = 0$ and which is holomorphic in ρ for all ω's in the interval $(0, 2\pi)$. Now, the general integral of (11) is

$$\frac{1}{\rho} + \log \rho - K\omega = \text{const.},$$

and, therefore, a solution ζ of (11a) assuming the same values as ρ when $\omega = 0$ will evidently satisfy the relation

$$\frac{1}{\rho} + \log \rho - K\omega = \frac{1}{\zeta} + \log \zeta. \tag{12}$$

It is thus seen that the required integral will be

$$\mathcal{J} = \zeta^2,$$

ζ being defined by (12).

Now, the question whether ζ and, consequently, $\mathcal{J}$ are holomorphic in ρ for all sufficiently small values of the latter variable is not, at first sight, as simple as it usually is in similar problems: for the common value of both sides of (12) will become infinite, and even in a rather complicated way, for $\rho = 0$. However, in order to resolve this question, or the more general one, which Poincaré does, for the relation

$$\frac{1}{\rho} + \beta \log \rho + Q(\rho) - K\omega = \frac{1}{\zeta} + \beta \log \zeta + Q(\zeta) = V, \tag{12'}$$

with $Q(\rho)$ holomorphic in ρ [to which we should be led by integrating similarly to (11) any analytic equation of the type

$$\frac{d\rho}{Kd\omega} = P(\rho) = \rho^2+\beta\rho^3+\gamma\rho^6+\cdots],$$

we need only to change variables by writing

$$\zeta = \rho(1+u),$$

by which (12) becomes

$$\frac{1}{\rho}+\log\rho - K\omega = \frac{1}{\rho}(1-u+u^2-\cdots)+\log\rho+u-\frac{u^2}{2}+\cdots,$$

or

$$u-u^2+\cdots-\rho\left(u-\frac{u^2}{2}+\cdots\right)=K\rho\omega, \tag{13}$$

the term $\rho Q[(1+u)\rho]-\rho Q(\rho)$ having to be added if we are concerned with (12′). Now, (13) will admit, for u, a solution represented by a power series in ρ with a radius of convergence bounded below for every finite value of ω and especially for $0\leq\omega\leq 2\pi$, as is immediately shown by the theorem of implicit functions.

Poincaré himself does not operate exactly as we have just done: he uses two different methods, one of which consists in introducing u not in (12), but in the differential equation

$$\frac{d\zeta}{P(\zeta)} = \frac{d\rho}{P(\rho)}$$

between ρ and ζ, the transformation thus obtained belonging to the class studied in the well known work of Briot and Bouquet.

Another proof rests on a direct study of the function ζ defined by (12′). We must confess that it does not seem to

us to be entirely convincing;[1] but this is not important, as our first proof is rigorous.

The quantity u being thus calculated, we go on as usual in the "Calcul des limites," by noticing that

$$\mathcal{J} = \zeta^2 = \rho^2(1+u)^2$$

is necessarily the result which we should obtain if we treated the partial differential equation (11a) in the way described above (every integration with respect to ω being carried out from the lower limit zero). The corresponding power series in ρ is convergent, and so is, *a fortiori* the analogous one obtained by dealing in the same way with (9a).

On the other hand, we know that the latter expansion, under our assumptions,—i.e., if (9a) is obtained from our original equation (1)—will be entirely trigonometric in ω. Such an expansion does not yet give us a satisfactory answer to our question because it does not satisfy the conditions that the coefficient of ρ^q only introduces multiples of ω with the same parity as q; or, in other terms, it does not remain invariant by changing ρ into $-\rho$ and ω into $\omega+\pi$. But this difficulty is easily solved: for such an invariance belongs to the quantity Ψ which intervenes in the partial differential equation (9a), so that, $F(\rho,\omega)$ being a solution of that equation, another solution will be

$$F(-\rho, \omega+\pi),$$

and a third one

$$\tfrac{1}{2}[F(\rho, \omega)+F(-\rho, \omega+\pi)].$$

[1]Poincaré writes (*Jour. de Math.*, 4th ser., I, 187): "Il faut démontrer ensuite que ζ revient à la même valeur quand ρ décrit dans son plan, et ω dans le sien, un contour suffisamment petit enveloppant le point zéro. Or, dans ces conditions, le premier et, par conséquent, le second membre de l'équation augmenteront d'un multiple de $2i\pi$, ce qui fera décrire au point ζ un contour fermé enveloppant le point zéro." This would be valid if ζ were a uniform function of V, the common value of the two sides of the equation: which he does not prove.

This one possesses the invariance in question and satisfies, therefore, all the conditions of the problem, which was to be proved.

We see therefore that in the above treated case of X,Y holomorphic in (x,y) about the origin, the vanishing of all the C_0's is a necessary and sufficient condition in order that this point be a center.[1]

The difficulty will be, in general, to verify the infinite number of conditions thus obtained. Poincaré points out a simple case where their existence can be asserted *a priori*, viz., when the differential equation is symmetrical with respect to the axis of x, which precisely occurs in some problems suggested by Celestial Mechanics.

Let me say that this is one of the first occasions on which Poincaré introduces stability questions; stability being said to take place in the case of a center, where every trajectory is closed, at least inside a certain region R. Poincaré shows that if R does not include the whole plane, its boundary necessarily goes through a singular point of the equation. On the contrary, the case of a focus, where trajectories indefinitely approach the origin or more and more recede

[1]Bendixson, in his Memoir of the *Acta Math.*, XXIV (see below) solved the question in another and quite direct way. He does not start from the partial differential equation, but from the original one, written in polar coordinates

$$\frac{d\rho}{d\omega}=h_2\rho^2+h_3\rho^3+\cdots.$$

If we again suppose that the data are holomorphic, ρ will be (see below) a holomorphic function of ω and ρ_0, the latter denoting the initial value of ρ for $\omega=0$. As, moreover, ρ must vanish with ρ_0, we shall have

$$\rho=\rho_0+k_2(\omega)\rho_0^2.$$

Now, every successive k will be determined by an integration with respect to ω, in terms of the preceding one and the h's. Things will behave as in Poincaré's discussion, the question being, each time, whether k will be of a trigonometric form or, on the contrary, will contain a secular term. In this method, there is no difficulty as to convergence, ρ being known *a priori* to be holomorphic with respect to ρ_0.

from it until they leave the region where expansions in powers of x and y are (in general) available, corresponds to instability.

But Poincaré also deals with stability or instability as concerns trajectories not contained in the immediate neighborhood of one point by choosing proper instances of equations (of the first order, but not of the first degree) belonging to the type

$$\frac{d\rho}{d\omega} = f(\rho).$$

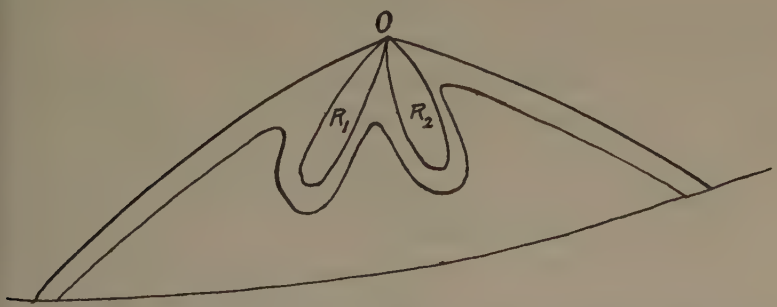
This study interests us as having led him, for the first time, to the consideration of the Poisson stability, which will play an important part in his further researches: as he finds, it may happen that a trajectory, issuing from any given point P, will return an infinity of times to an infinitely close neighborhood of P (though, perhaps, after having gone, meanwhile, farther and farther away from it, to more and more remote regions of the plane), which is the definition of Poisson stability.

This study of singular points of a differential equation has been pursued by several authors[1] in the direction of higher singularities. These treat the problem rather from the analytic point of view, in the complex field. Such is not the case with Bendixson,[2] especially in his paper in the *Acta Mathematica*, XXIV (1900). Bendixson arrives at

[1]Besides Picard (*Traité d'Analyse*, III, 2 ed., 23–30; *Comptes Rendus*, LXXXVII, 1878, 430, 734; *Bull. Soc. Math., France*, XII, 1883–1884, 48), we have: J. Horn, *Zeitschr. Math. Phys.*, 49 (1903), 246; *Arch. Math. Phys.*, ser. 3, VIII (1905), 237. E. Lindelöf, *Acta Soc. Sci. Fennicae*, 22 (1897), Mém. n° 7, 1–26. H. Dulac, *Jour. Ecole Polytech.*, ser. 2, ch. 9 (1904), 1–125; *Ann. Univ. Grenoble*, 17 (1905), 1; *Jour. Math. pures et appliquées*, ser. 6, II (1906), 381; *Bull. Soc. Math., France*, ser. 2, XXXII (1908), 230–252.

[2]We also mention a paper of O. Perron (*Math. Zeitschrift*) where Poincaré's theory is resumed with the minimum number of hypotheses concerning the functions X, Y.

general theorems valid for any sufficiently regular form of the functions X, Y, even including, at least in the first part of his Memoir, non-analytic ones. The conclusions show a remarkable generalization of what happens in the first case where X and Y begin with linear terms. If the singular point O is not a focus, there exist a certain number of trajectories abutting at O, which divide the vicinity of O into several regions, in some of which the differential equation behaves as it would in the neighborhood of a pass, while in others,—which are called *nodal regions*—things take place as around a node, i.e., every trajectory from a point within such a region abuts on O. It may happen that trajectories of the latter kind end at O only for $t = +\infty$ or only for $t = -\infty$, such nodal regions being then called *open* ones; but we may also find *closed* nodal regions, which are such that every trajectory through them ends at O as well for $t = +\infty$ as for $t = -\infty$, the trajectories enclosing each other as shown in the diagram below.



Regions of this kind, if they do not include the whole surface, must have for their boundary a trajectory which contains another singular point, as is the case for the region around a center where trajectories are closed.

The only circumstance showing a notable difference from the previous discussion is the possibility of a pass-like region

including in its inside one or several nodal regions R_1, R_2, $\cdots$, the trajectories within such a pass-like region behaving as shown in the diagram on the preceding page.

The reader of this paper will notice a difference between Poincaré's and Bendixson's definitions of centers: for a singular point O is called by Bendixson a center if there is an *infinity* of closed trajectories surrounding O and indefinitely near it, while, in Poincaré's work, a point is a center only if *every* trajectory passing within a sufficiently close neighborhood of O is closed. But this is an instance in which analytic equations and non-analytic ones may be profoundly different. If X, Y are holomorphic around O, either all trajectories passing sufficiently near O are closed or none of them is.

This results from the fundamental theorems, completed by Poincaré (as we have previously seen)[1] by the study of the influence of initial conditions. Let us take our given equation in polar coordinates, with ω as the independent variable, with respect to which it will behave regularly. Then, ρ_0 denoting the initial value of ρ for $\omega = 0$, the value ρ_1 assumed by ρ for another given value ω_1 of ω will be a holomorphic function[2] of ρ_0 around $\rho_0 = 0$ (vanishing with ρ_0). Let us take especially $\omega_1 = 2\pi$: the difference $\rho_1 - \rho_0$, being represented by a convergent power series in ρ_0, will either be identically zero or vanish for no ρ_0 between zero and a certain positive value.

A second part of the same paper is more especially devoted to the analytic case. In order to see in what direction

[1] See my previous Lectures, p. 176.

[2] This is, in the first place, proved inside a certain interval of values of ω around $\omega = 0$ or, more generally, around ω_0, if we should take an initial value ω_0 other than zero. But classical arguments show that the amplitude on such an interval has a lower limit independent of ω_0, so that only a finite number of them is required in order to cover the interval $(0, 2\pi)$: therefore, successive applications of the principle of analytic continuation in a finite number will give the conclusion in the text.

Poincaré's original discussion has to be generalized for higher singularities, let us mention the precise result obtained by Bendixson as to "nodal" regions. He succeeds in obtaining an upper limit for their number if X and Y begin with terms of order m; there cannot be more than $2m$ nodal regions around O.

We must not leave the subject of equations of the first order and the first degree without reverting to the first part of the theory as contained in the first two parts of Poincaré's Memoir and spoken of in my previous Lectures: indeed, we must mention that this part of this theory has been given, in recent times, a quite new and remarkable development. On account of the very boldness and profoundness of Poincaré's ideas, it could be imagined that he worked for the far future. In fact, some time elapsed before the various parts of his work could be continued, though this was done in several, but always theoretical, directions. Now, it is the more remarkable that, for some years, the methods in the Memoir *Sur les courbes définies par une équation différentielle* have been applied in a quite effective and practical way, for engineering purposes. The question arises on account of the fact that, when treating real and practical dynamical problems, it is absolutely impossible to abstract from dissipative forces, and these completely upset the classic theory of oscillations around a position of equilibrium. Periodic motions occur under quite other circumstances than those classically known, namely when there are auxiliary forces which depend on speed, so that the equations can no longer be assumed to be linear.

The very curious phenomena thus arising and which theory as well as experiment show to be very different from our previous ideas, were, for the first time, discovered by

Van der Pol;[1] these have been studied rather thoroughly as well for natural as for forced oscillations. Now, such a study, especially in the latter case, depends on Poincaré's methods, the question being to investigate the limiting cycles; and Poincaré's principles have been used for actual, numerical calculation.[2]

[1]See *Phil. Mag.*, 7 ser., II (1926), 978; *Jahrb. der drahtlosen Telegr. und Telephonie*, XXVIII (1926), 178; etc.

[2]See papers of Cartan, Lienard, Le Corbeiller, and, more recently (in the *Archiv für Elektrotechnik*), Andronon.

III

EQUATIONS OF HIGHER DEGREE AND THE INTERVENTION OF ANALYSIS SITUS

AFTER this discussion of centers we come back with Poincaré to the general equation of the first order

$$f(x, y, y') = 0. \tag{1}$$

If f is of the first degree in y', we have the problem already treated. We have seen that the most essential feature of the results obtained is the existence, in general, of limiting cycles: every trajectory which does not end at a singular point, if not closed, is asymptotic to a closed trajectory or "limiting cycle" (and we can even consider a certain class of singular points, namely, foci, as a special kind of limiting cycles). It would appear at first sight that if f is of higher degree in $z = y'$, we must have something entirely different, but this is not the case. In a very large class of cases, the new problem is not different from the old. It is solved in the same way and by use of the same principles; the character of the solution does not depend upon the degree of the surface either with respect to z alone, or to x,y, and z. It depends upon the properties of the surface, represented by equation (1), with respect to Analysis Situs.

I hope I need not dwell long on what Analysis Situs is. It is concerned with properties which remain unchanged when arbitrary continuous transformations are made. Two surfaces (or geometric beings of any kind) which by continuous smoothing, without any tearing or fastening (think

of them as made of wax), can be transformed into each other, are said to be equivalent from the point of Analysis Situs, or, after Poincaré's terminology, homeomorphic.

Now, it is this point of view of Analysis Situs which proves to be fundamental in the study of differential equations.

This new and essential idea again bears that character of marvelous simplicity which we had to point out, in our first series of lectures. Thinking it over from our present point of view, now that Poincaré has actually taught us how to consider things, we should be tempted to wonder, not how he has been able to discover this rôle of Analysis Situs, but how geometers have had to wait on him for it, so necessary and unavoidable its introduction now appears to us to be.

Let us remember the main features of the history of Analysis Situs. You know that in the time of Euler it may, at least at first sight, have been hardly more than an occasion for rather childish problems, such as the 36 officers, the bridges of Königsberg, and so on. However, there is already one important exception, namely, Euler's classic theorem on polyhedrons, expressed by the formula

$$F + S = A + 2, \tag{1}$$

where F, S and A respectively mean the numbers of faces, vertices, and edges of an arbitrary polyhedron.

Now, this very example is connected with an especially curious historical fact, one of the most interesting for every one who (I must confess that such is my case) cares not for the history of Science in itself, but for what we have to learn from it, for those instances in which the past can teach us how to act or how not to act in the future. You know that Cauchy began his scientific work with a Memoir

(1813) devoted to pure and even elementary geometry and, precisely, to properties of polyhedrons. He gives, or thinks to give, a proof of Euler's theorem (1), which proof has been accepted and taught for years.[1] The proof intends to be absolutely general, not assuming convexity or any other special property of the polyhedron: therefore, of course, it must be and is false.

More than thirty years later, Cauchy had the honor of founding the theory of analytic functions, while Riemann was doing the same thing independently on his side. Though both discoverers attained the same general scope, in several respects our debts to them are unequal. In particular, Cauchy's theory could not have been sufficient to build up the theory of algebraic and abelian functions: one essential element, the genus, would not easily have been set in evidence, whereas, with Riemann, it appears quite intuitively from the most elementary and direct considerations of Analysis Situs. I cannot help thinking that, if Cauchy had perceived and corrected his error in 1813, he would have added this gem to his crown in 1850, instead of leaving it to his rival.

Now, the intervention of Analysis Situs in this case is really a quite general fact, the bearing of which goes far beyond the domain of algebraic or even analytic functions. Its origin is the very essence of Integral Calculus. Riemann's is the first instance in which Integral Calculus—that is, the inference of finite properties of a variety from infinitesimal ones—proves insufficient when left to its own resources exclusively. If we are given only the latter properties, something may, and will very frequently, be lacking

[1]The authors of one of the most popular Treatises (at the time when I was a student) apparently felt some trouble with it, as they most candidly showed by literally reproducing Cauchy's own text within quotation marks, thus leaving him the entire responsibility for it.

in our knowledge of the former, and this is supplied only by topological elements. These may even be completely independent of the properly analytic ones (which is not the case with algebraic functions). Between ordinary Euclidian space as we are accustomed to conceive it and Klein-Clifford's periodic space, there is no difference whatever in a sufficiently restricted domain: we do not know that this space has not actually the Klein-Clifford structure, and we never shall, because *local* properties of both spaces are rigorously identical.[1]

Riemann's example ought, as we see, to have led geometers to foresee the introduction of Analysis Situs into problems of Integral Calculus, therefore into the integration of differential equations also, as an absolutely necessary thing. Nevertheless, this was not perceived until Poincaré came.

On the other hand, Analysis Situs is not to be introduced at once: another precaution is necessary, the conception of which is also due to Riemann. In fact, Riemann had to correct and improve the very definition of a function of a curve. It is by a misuse of language that we are accustomed to say that the ordinate y of a circle is a function of the abscissa x: this would be true only if every given value of x gave one determinate value of y, whereas it actually gives two. From this point of view (as we already noticed in our preceding Lectures), we ought to say that a non-uniform function of x is not truly a function of x, as does Riemann, by considering such a function y as dependent not on the

[1]A mechanical model of a Klein-Clifford space can be constructed very easily: let us describe it, for simplicity's sake, in two dimensions only. Let a vertical material axis bear a rigid piece of an arbitrary shape which can revolve freely about it and which, in its turn, shall bear the axis AB of a rigid homogeneous body of revolution which can also rotate freely around AB. If each of these two independent motions is assumed to be frictionless, such a system will be governed, *locally*, by the same equations as a material point moving freely and frictionlessly on a horizontal plane. Nevertheless, the finite properties of the two motions will be utterly different, since, in the former, the trajectories do not go to infinity.

location of a point in a plane, but on the location of a point on a several-sheeted surface, on which, and only on which y can be said to be well defined.

An elementary and, for that very reason, a curious instance of the importance of this is given by the question of singular solutions. It is well known that, for our equation (1), the singular solution is obtained as the envelope of the general integral: and, on the other hand, it is classic that we obtain such an envelope by eliminating y' between (1) and

$$\frac{\partial f}{\partial y'} = 0. \tag{1'}$$

Consequently, it has been taught for long years that equations of the form (1) in general admit of singular integrals.

The conclusion is false. The fault lies in the fact of thinking that our equation correctly defines y' as a function of (x,y), while (1), for each system of values of x,y, gives *several* values of y' and, moreover, these cease to be regular functions of x,y precisely when (1) is satisfied, as the theorem of implicit functions fails then to be valid. Geometrically speaking, there will be several vectors at each point which represents (x,y) in the plane or on the sphere. If we desire a vectorial distribution which is well defined (singular points excepted), we are not justified in considering it on a plane. We must consider it on the surface

$$f(x, y, z) = 0. \tag{2}$$

If, at M, a point of the surface, we draw the plane $dy/dx = z$, the intersection of this plane with the tangent plane gives a definite direction and we are thus provided with a field of vectors tangent to the surface, with the condition that there is a single vector at each ordinary point.

Now, and only now—that is, when the latter condition, which is fundamental, is satisfied—we can give a correct answer to the question of what happens at any point where we have the relation (1′).

Geometrically, this equation defines the apparent contour of the surface (2) with respect to the xy-plane, that is, expresses that the tangent plane to the surface is parallel to the z-axis. But the plane $dy/dx = z$ is also parallel to the z-axis: so will be, therefore, their intersection, which is the vectorial direction associated with our point, if this direction is well determined.

If so, our point will be an ordinary point for the vectorial distribution and, consequently, for our differential equation or, more exactly, for the geometrical problem corresponding to it, *when considered on our surface* (2): through such a point, we shall have, in space, a perfectly determined curve satisfying the differential equation. But let us now come back to the ordinary conception of the problem, that is, project the whole diagram onto the xy-plane. The above curve, being projected onto a plane parallel with one of its tangents, will, as is well known, give a cusp in its projection (with a cuspidal tangent which will depend on the osculating plane to our curve in space and, in general, not lie in the tangent plane to the surface); and this is precisely the result which Darboux had proved by a special calculation, and which now becomes quite intuitive.

The only exceptional case occurs when the two planes coincide, namely,

$$\frac{\partial f}{\partial x} + z\frac{\partial f}{\partial y} = 0, \tag{2′}$$

so that they no longer determine the vectorial direction. If this takes place at one point of the apparent contour, we have a singular point as previously examined and,

therefore, *on the surface*, a singular point of Poincaré's diagram: in general, a node, a focus or a pass: but, on the xy-plane, shapes will be different on account of the singularity of the projection.[1]

Evidently, we must proceed in the same way for any general study of differential equations: such an equation must not be considered as defining a curve on the xy-plane, but a curve on the surface (2)—namely, a curve such that $dy/dx = z$.

It may happen that this surface (2) is of genus zero, that is, topologically equivalent to a plane. Then, every result of the former theory, such as spoken of in my Lectures of 1920, will apply. For these results depend on two principles: 1° that two trajectories cannot intersect, except at a singular point; and 2° that a closed curve separates the surface into two pieces. The first principle is valid for any surface; the second is true for all surfaces which can be continuously transformed into a sphere (surfaces of genus zero). Thus, the theory in the first two parts of the Memoir is much more general than appears at first glance. It applies to a great many equations of the first order and higher than the first degree.

I shall not follow Poincaré throughout his researches concerning higher genera. The study is up to a certain point similar to that of the earlier case, but there are certain modifications. There is one case, however, in which the results are new and especially interesting: I refer to the surfaces of genus one, such as the anchor ring.

[1]Walther Dyck has discussed such singularities at points satisfying (1′). As we see, the results can be foreseen at once, by projecting Poincaré's diagrams onto our horizontal plane: however complicated the new diagrams may look at first glance, the only difference from the original ones arises from the fact that every meeting point of the curve with the apparent contour gives a cusp on the projection.

In various mathematical theories dependent upon Analysis Situs we often have to divide our surfaces into two main classes, those of genus $0, 2, 3, 4, \cdots$, and those of genus 1, the case of the genus 1 being special. At bottom this is connected with Euler's theorem. If we think of the surface as made up of (in general, curvilinear) polygons put together, we have, for a surface of genus 0,

$$F+S=A+2,$$

where F, S, and A are the number of faces, vertices, and edges respectively. If the genus is p, the relation

$$F+S=A+2-2p$$

holds; and if $p=1$ it becomes

$$F+S=A.$$

Poincaré points out the importance of this in the relation concerning the singular points. We again define, at each point, an index number, which is 0 at an ordinary point, which is 1 for a pass, and -1 for a node or focus. The index number is essentially a measure of the rotation of the vector as a small closed curve about the point is traced. On summing the index numbers for the whole surface and employing Euler's theorem, we find

$$c-f-n=A-F-S=2p-2,$$

where c, f, n represent the number of passes, foci, and nodes, respectively. Hence, for any vectorial distribution tangent to the surface, there must be singular points, except in the case $p=1$.

The continuation of the Memoir *Sur les courbes définies par une équation différentielle*, more exactly of the third part of it, deals precisely with the hypothesis that there is no singular point, the surface (2) being of genus 1 and,

therefore, reducible, without loss of generality, to a torus. We now know that no other value of the genus would be compatible with that hypothesis.

But, before reviewing this study of the torus, we must glance at other works to which Poincaré was almost unavoidably led by the above course of ideas. Knowing, as we now do, the importance of Analysis Situs, and that every passage from local to general points of view, that is, every process of integration, may be profoundly influenced by it, we must conclude that improvements in that branch of science must become, one day or another, essential to Integral Calculus: and that, just as we have been obliged to introduce orders of connection in the study of the general differential equation of the first order, we must expect to be unable to treat correctly differential systems or equations of higher order without a knowledge, as thorough as possible, of Analysis Situs in more than two dimensions. We must not wonder, therefore, that Poincaré was compelled to follow that course and that Lefschetz was able to begin his Treatise on Topology[1] with the following lines:

"Perhaps on no branch of Mathematics did Poincaré lay his stamp more indelibly than on Topology."

On quoting this statement of Lefschetz, let us say immediately that this subject of Analysis Situs has been and continues to be most brilliantly cultivated on American soil. The names of Veblen, Alexander, Lefschetz will remain attached to the foundations of n-dimensional Topology, among those of the most prominent successors of Poincaré in that line. For there was much to be done, even after Poincaré's powerful impulse: there is much to be done even now. The subject is one of the most difficult in present mathematical science.

[1] *Am. Math. Soc. Colloquium Publications*, XII (1930), New York.

Let me remind you that it now contains two very distinct parts. A considerable series of profound results, which are topological in the sense of the above definition, depend on the modern theory of sets and even represent the most important geometrical applications of that theory. This branch of Topology, which has been considerably developed in recent times, begins, for instance, with the proof of Jordan's theorem. On the other hand, such theorems as Jordan's are considered as evident and admitted once and for all in Topology as Riemann understood it, that is, in what is now called Combinatory Topology.

It is this latter kind of Topology which was undertaken by Poincaré; but, even when the question is reduced to these terms, there is no comparison, as to difficulty, between the case of higher numbers of dimensions and the one which was solved by Riemann's principles. Poincaré himself, and others as well, made some mistakes. For example, Poincaré stated that there could not be a closed "one-sided" surface. Yet, the following figure, similar to a sort of fly-trap used in France, is an instance (due to Klein) of a surface of such a kind. But the mathematician does not need to construct artificially such examples, as there is one which is quite classic, no other than the projective plane.

We shall only mention some of the first and simplest discoveries of Poincaré in Topology, the more because, as we have said, the subject is a most familiar one in America.

To begin, he had to get a quite new view of what is a variety, or of what is a boundary, not excluding the possibility of counting some edges, faces, etc., several times in a boundary. It is in this sense that between several closed $(n-1)$-dimensional varieties $v_1, v_2, \cdots$, in an n-dimensional one V, he considers "homologies," consisting in the fact that proper combinations of the v's constitute the complete boundary of a part of V.

It is well known—and this is practically all that was known before Poincaré—that the analogue of Riemann's order of connection, for n-dimensional varieties, is given by Betti numbers, of which there are $n-1$ (that is, for $n=3$, the "linear connection" and the "superficial connection"). Analogy with Riemann's case inclined one to believe that the knowledge of these Betti numbers, together with the number of boundaries, was sufficient to characterize the variety from the topological point of view, so that, for instance, two closed varieties with the same Betti numbers were necessarily homeomorphic. One of Poincaré's first tasks was to show that this idea was false; and, as we shall see below, further discoveries had shown more and more how complicated things are from this point of view.

Between the numbers of Betti, Poincaré discovered a beautiful relation, namely, that they are equal two by two the first one being equal to the last, the second to the last but one, and so on. Poincaré proved this by a proper application of Kronecker's index[1] to intersections of k-dimensional and $(n-k)$-dimensional varieties in V; it has been noticed, since, that such a relation can be made almost intuitive by the introduction of a topological duality in polyhedral varieties. Such a duality can be understood by its simplest instance, which is given by the classic polygonal

[1]See my Lectures of 1920, p. 150, 163.

divisions on the sphere corresponding to two conjugate regular polyhedra, each polygon of the first division being such that one point inside it (in this case its center) coincides with a vertex of the other division, and conversely.

These Betti numbers P_i intervene in the n-dimensional generalization of Euler's known theorem on polyhedra, the right-hand member of the equation which expresses such a generalization being $P_1 - P_2 + P_3 \cdots$. One curious consequence of this is that, for odd n's, this right-hand member and, therefore, the left-hand one, vanish, on account of the equalities between the P's.

But, notwithstanding the simplicity of the relations between Betti numbers, their proof by Poincaré admitted of an objection which was pointed out by Heggaard, after which Poincaré returned to the subject in five new Memoirs.[1] The difficulty consisted in the fact that it is not allowed to divide a homology by an integer, even if the latter appears as a factor in all the coefficients. Now, this gave place to the discovery, by Poincaré, of new topological invariants, the "torsion numbers," so that, in order that two varieties be homeomorphic, it is necessary that not only their Betti numbers, but also their torsion numbers be the same.

Even these new conditions were not sufficient, and Poincaré introduced a new element which is connected with the topological character of the variety, namely, a group. To construct this group, he introduces certain functions, F_1, F_2, $\cdots$, defined by a system of total differential equations

$$dF_1 = A_{11}dx_1 + \cdots + A_{n1}dx_n, \text{ etc.},$$

where the A's are functions of the coordinates and of the F's, and satisfy the conditions for total integrability.

[1] *Rend. di Palermo*, XIII; *Proc. London Math. Soc.*, XXXII; *Bull. Soc. Math., France*, XXX; *Jour. de Math.*, ser. 5, VIII; *Rend. di Palermo*, XVIII. The first Memoir had been published in the *Jour. École Polytech.*, ser. 2, I.

However, if the region is not simply connected, we get, as a result of following closed circuits, a discontinuous group of transformations on the F's. Poincaré shows that to each variety there corresponds such a group; and two varieties cannot be homeomorphic if the corresponding groups are different.

Again, it might be thought that this new character, combined with the former ones was sufficient to give a complete topological description of the variety. But Alexander has shown that such is not the case, so that, on this question, the final word has not yet been said.

We thus see how abstruse the matter is. It has not ceased to give place to the works of the American school, to which we must add the names of Hermann Weyl and another young geometer, Heinz Hopf. One of the most interesting results of this recent work has been to connect the topological nature of the variety—more precisely, the right-hand member of the formula which generalizes Euler's theorem—with the distribution of singular points of any vectorial distribution over the variety: a fact which will doubtless be of importance with respect to differential equations, as is shown by the instance of the two-dimensional case.

IV

THE CASE OF THE TORUS

WE shall now make a study of the integral curves on the torus under the hypothesis that there are no singular points whatever. We shall use the coordinates φ and ω, where φ is the longitude, or azimuth angle, and ω is a latitude on a meridian circle, each value of ω defining a determinate parallel on the surface. With Poincaré, we shall agree not to measure the angles φ and ω in the ordinary way, but with the whole circumference, that is, 2π radians, as unit: in other words, to divide the ordinary measure by 2π. Thus the simultaneous knowledge of φ and ω defines a quite determinate point on it. Moreover, this point does not change if we increase either φ or ω by any integer n. However, it is important to notice that, when we follow continuously any trajectory on the surface, the final values of φ and ω at any point of that curve will be perfectly defined if such values are given at an initial point.[1]

Our differential equations will be given in the form

$$\frac{d\varphi}{dt} = \Phi, \quad \frac{d\omega}{dt} = \Omega \tag{1}$$

(Φ and Ω being given functions of φ and ω, which are continuous, periodic, and never vanish simultaneously).

[1]It may also be convenient, sometimes, to consider ω' as equivalent to ω only after several revolutions: for example, to call ω, $\omega+3$, $\omega+6$, etc., equivalent [$\omega' \equiv \omega \pmod 3$]. This amounts to considering a sort of Riemann surface on the torus, the revolution of a point being complete only when it has turned three times; and the same can be done as to φ.

To avoid superficial and uninteresting difficulties, we shall assume that Φ is constantly positive, so that φ is constantly increasing; and we can even admit that Ω is always positive and ω constantly increasing.[1]

Now, starting with a solution curve from $\varphi=0$, $\omega=\omega_0$, let φ increase continuously. Then ω varies continuously. When $\varphi=1, 2, 3, \cdots$, let $\omega=\omega_1, \omega_2, \omega_3, \cdots$. We shall speak of ω_1 as the first consequent of ω_0, ω_2 as the second one, and so on. Letting φ decrease, we get likewise the antecedents. As we have said, the values of the ω's are fully determinate when ω_0 is given, say

$$\omega_1=\psi(\omega_0), \tag{2}$$

so that, to begin, we can plot them on a straight line. If (as assumed) $\Omega>0$, then we have $\omega_0<\omega_1<\omega_2\cdots$.

Our fundamental hypothesis is that two trajectories never intersect. It follows from this that:

1° $\psi(\omega)$ is a constantly increasing function of ω;

2° when the initial point describes once the circumference of a meridian, its consequent will also describe once, and only once, the same circumference, so that

$$\psi(\omega_0+1)=\psi(\omega_0)+1.$$

The first property can be expressed in two other forms, both of which will be of special interest for us, namely,

1ª $\psi(\omega)$ gives a bi-univocal and bi-continuous correspondence between ω_0 and ω_1 and, likewise, there is a bi-univocal and bi-continuous correspondence between ω_0 and ω_i.

[1]That this second hypothesis is no real restriction of generality appears immediately by considering, instead of (1), the equations

$$\frac{d\varphi}{dt}=\Phi, \quad \frac{d\omega}{dt}=\Omega+l\Phi$$

(that is, changing ω into $\omega+l\varphi$), where l is a constant—say an integer—which can be chosen so that $\Omega+l\Phi$ is always positive.

1^b Let us replace ω_0 by another initial value $\bar{\omega}_0$, and let us suppose that the latter lies, say, between ω_0 and ω_1. Then, $\bar{\omega}_1$ will lie between ω_1 and ω_2; and, going on in the same way, we see that *the $\bar{\omega}$'s and the ω's will separate each other.*

We have to deal with $\psi^{(2)}(\omega_0) = \psi[\psi(\omega_0)]$, $\psi^{(3)}(\omega_0) = \psi[\psi^2(\omega_0)]$, etc., that is, with the iteration of the function ψ. The principle 1^b just enunciated, shows us that, from the qualitative point of view, the result will be, to a certain extent, independent of the value of ω_0.

Finally, we shall have to take account of the fact that the successive iterates can be augmented or diminished by any integer without changing the geometrical result, so that only the fractional parts ω_i' or, as we shall say, the "reduced values" of the successive quantities ω_i will finally interest us. Thus, the differential equation has disappeared from view and the problem is now of an arithmetical character.[1]

But, for the present, let us still consider the exact values of the ω_i's (and not only their reduced values). A fundamental result is found by Poincaré, namely, that there exists a limit

$$\mu = \lim \frac{k}{j(k)}, \tag{3}$$

$j(k)$ being the number of consequents of ω_0 found within k revolutions around the meridian circle, and, moreover, this limit μ does not depend on the value of ω_0.

In an important Memoir,[2] to which we shall have to return at greater length, Birkhoff has given the proof in

[1] It may even happen that properly arithmetical researches, such as undertaken in recent times chiefly in connection with Riemann's theory of prime numbers, will help in solving the remaining questions in the present study.

[2] *Acta Math.*, XLIII (1920).

a simple and elegant form. Let ω_i be the i^{th} consequent of ω and let us consider the difference $\omega_i - \omega$. If ω, starting from any determinate value ω_0, increases to $\omega_0 + 1$, then its i^{th} consequent will increase from ω_i to $\omega_i + 1$, so that the difference in question will come back to its original value, having therefore a minimum a_i and a maximum b_i. On the other hand, it can never increase by 1, so that

$$b_i - a_i < 1, \text{ and } \frac{b_i}{i} - \frac{a_i}{i} < \frac{1}{i}.$$

Let us now replace the integer i by another integer j, so that a_i, b_i are replaced by a_j, b_j. The two intervals

$$\left(\frac{a_i}{i}, \frac{b_i}{i}\right) \text{ and } \left(\frac{a_j}{j}, \frac{b_j}{j}\right) \tag{4}$$

must have a common part. This results from the consideration of the difference $\omega_{ij} - \omega$, which must lie between ja_i and jb_i, and also between ia_j and ib_j.

This being the case for any two integers i and j, we see that every a_i/i is less than any b_j/j. As, on the other hand, the intervals (4) are infinitesimal when i or j become infinite, we see that, under these conditions, they approach a common limit μ and this is the required limit, $2\pi\mu$ being the "rotation number" (according to Birkhoff's terminology).

We also see what must take place if μ is commensurable, say $\mu = p/q$: then, if we take $i = q$, we see that p is contained between a_q and b_q. Therefore, in any interval of amplitude p there must lie at least one and even two ω_0's such that $\omega_q - \omega_0 = p$: therefore, two closed trajectories.

Moreover, p/q is contained within any one of the intervals (4).

Conversely, if there is a closed trajectory, there must be an ω_0 such that $\omega_q - \omega_0 = p$, where p and q denote two

integers; we shall have also $\omega_{nq}-\omega_0=np$ for every integer n, and, therefore, the limit of $(\omega_N-\omega_0)/N$ cannot be other than p/q. Thus, the commensurability of μ is a necessary and sufficient condition for the existence of closed trajectories.

Poincaré reaches the same results by directly considering (3). Let us assume, for instance, that there is no closed trajectory, and let us denote, as above, by $j=j(k)$ the number of consequents of ω_0 contained within k revolutions of the meridian circle (starting from ω_0), ω_0 itself being included. Under our present hypothesis, ω_j must lie beyond ω_0+k (and not coincide with ω_0+k). Then, as we know that j successive consequents of any point will be separated by the points having arguments $\omega_1, \cdots \omega_j$, we see that any k consecutive revolutions will contain either j or $j-1$ such points. Moreover, such a sequence of k consecutive revolutions (whatever be their origin) will contain either j or $j-1$ consequents of ω_0. Let now l be another integer: the number of consequents of ω_0 contained within the first kl revolutions will be contained between $l[j(k)-1]$ and $lj(k)$. But, similarly, it must be contained between $k[j(l)-1]$ and $kj(l)$, so that any two intervals such as

$$\left(\frac{j(k)-1}{k}, \frac{j(k)}{k}\right), \quad \left(\frac{j(l)-1}{l}, \frac{j(l)}{l}\right) \tag{5}$$

must have a common part. This again implies the existence of the limit $1/\mu$ which, moreover, must be contained in each interval such as (5).

Assuming now μ to be incommensurable, so that no closed trajectories exist, let us consider the circular order of the successive consequents: that is, we plot the values of the successive ω_j's on the meridian circle—and no longer on a straight line—or, what comes to the same thing, we reduce each of them to its fractional part ω'_j, or "reduced

value" by subtracting $E(\omega_j - \omega_0)$ where $E(u)$ denotes, as usual, the greatest integer contained in u. The successive consequent points M_j, under our present hypothesis that μ is incommensurable, will be infinite in number.

Now, under the same hypothesis of μ incommensurable, this integer $E(\omega_j - \omega_0)$ is in a direct relation with μ; for we know:

1° that $(\omega_j - \omega_0)/j$ is included between a_j and b_j;

2° that a_j and b_j never include any integer (else, by an earlier argument, there would be a closed trajectory);

3° that they include μ.

Then it appears that $E(\omega_j - \omega_0)$ is nothing else than $E(\mu j)$.

We can say more, and assert that *the circular order of the ω_j's* (which we already know to be independent of ω_0) *is the same as the circular order of the numbers* $\mu j - E(\mu j)$. This appears more easily from the first definition (3) of μ. First, $E(1/\mu)$ is nothing else than $j-1 = j(1)-1$. The j^{th} consequent will be, then, beyond the initial point M_0, but before M_1. Now, if, simultaneously, we consider, on another circle, the points $m_1, m_2, \cdots, m_{j(1)-1}$, the arguments of which (still measured with 2π radians as unit) are $\mu, 2\mu, \cdots, (j-1)\mu$, a j^{th} point m_j corresponding to the argument $j\mu$ will lie between m_0 and m_1, that is, the circular order m_0, m_j, m_1 will be the same as M_0, M_j, M_1: whence the analogous conclusion immediately ensues (on account of the above noted principle of separation 1[b]) as to the following points $M_{j+1}, \cdots$, contained within the second revolution.

Generally, let our conclusion be proved for the first k revolutions, the corresponding number $j(k)$ of consequents being equal to $E(k/\mu)$. Then, starting from our initial point in the positive sense, we shall successively find points

$M_0, M_i, M_{i'}, M_{i''}, \cdots$, the numbers $i, i', i'', \cdots$, being all less than $j(k)$: the circular order of these points is therefore, under our assumptions, the same as that of $m_0, m_i, m_{i'}, \cdots$; that is, the same as the circular order of $0, \mu i - E(\mu i), \mu i' - E(\mu i'), \cdots$. Moreover, $i, i', \cdots$, being not greater than $j(k)-1$, our conclusion is valid for the circular order of $M_{j(k)-1}$ with respect to $M_{i-1}, M_{i'-1}, \cdots$, and, therefore, for $M_{j(k)}$ compared to $M_i, M_{i'}, \cdots$; and, on the other hand, it holds also for $M_{j(k)}$ compared to M_0, on account of (5).

Holding for $M_{j(k)}$, it will, as above, hold for $M_{j(k)+1}, \cdots$, throughout the $(k+1)$th revolution; then, our last argument extends it to the $(k+2)$th one; and so on.

All this, evidently, suggests that we consider ω_i as a (constantly increasing) function of μi. This matter is connected, as we shall see, with the structure of the set P of all the consequents and antecedents of a given point.

If the trajectory from M_0 is not closed, these consequents and antecedents are infinite in number and, therefore, there will be a derived set P'. As is well known, such a derived set is closed, that is, it contains the second derived set P''.

The above argument shows us, under the same hypothesis that μ is incommensurable, that any arc between two points of P contains other points of the same set and, therefore, points of P'.

But, by another method, also contained in Poincaré's Memoir,[1] we can prove even more, namely, that if there is no closed trajectory, *any arc between two consequents*

[1]Poincaré applies the argument in the text only under the hypothesis that N belongs to P'. See p. 50 of this lecture. But it holds for any N, as we show in the text, and leads to the conclusions

1° that P' does not depend on the choice of M_0;

2° that P' is the same, whether deduced from the consequents of M_0 or from its antecedents.

M_i, M_{i+j} of M must include consequents or antecedents (and even an infinite number of consequents or antecedents) of any point N. For, let us admit that this arc M_iM_{i+j} would include only m of the consequents of N: then the arc $M_{i+j}M_{i+2j}$ would contain only $m' \leq m$ of them (the j^{th} antecedents of which would belong to the former); the same for $M_{i+2j}M_{i+3j}$, and so on. If a finite number of these successive arcs (which are all directed in the positive or all in the negative sense) covered the whole circle, N would have no consequent at all (which is absurd) or at most a finite number of them (so that the trajectory from N would be closed). But if they do not cover the whole circle, the successive points M_{i+j}, M_{i+2j}, $\cdots$, must tend to a determinate point γ, which, being together the limit of these points M and of their j^{th} consequents, must coincide with its own j^{th} consequent (on account of the fact that the function $\psi^{(j)}(\omega)$ is continuous) so that such a hypothesis implies the existence of a closed trajectory.

More precisely, we see that it implies the existence of a *limiting cycle*, such as we met with in the case of the sphere. Treating in the same way the antecedents instead of the consequents, we should find also a limiting cycle (perhaps coinciding with the former) for negative t's.

The above argument applies to the case where N coincides with M and gives, therefore, our former conclusion. But, in the contrary case, it immediately carries another one which is of evident importance, namely:

If there is no closed trajectory, *the derived set P' is independent of the choice of the initial point M.*

Moreover, the argument shows that any arc between two consequents or antecedents of M necessarily includes (if there is no closed trajectory) an infinite number of *antecedents* of M; and this settles an obvious question.

Evidently, the set P consists of two parts—the antecedents of M and the consequents—giving place each to a derived set. We see that these two derived sets coincide.

I do not intend to make any further study of the case of a limiting cycle or a closed trajectory;[1] and, therefore, I shall exclude it, as a rule, in what follows.

Now, let M be any point of P—for instance, M_0 itself. If M belongs to P', so do all its consequents (this again following from the fact that the solution is a continuous function of the initial conditions). Conversely, if M does not belong to P', then no point of P belongs to P'. We have, then, only two possible hypotheses: 1° that every point of P belongs to P'; and 2°, that no point of P belongs to P'.

If we refer to what has already been said (see p. 23) concerning stability, we see that the present distinction is equivalent to the distinction between stability and instability in the above sense—that is, Poisson's sense—1° corresponding to stability and 2° to instability. The latter hypothesis is evidently verified in the case which we have just studied, when there is a limiting cycle.

If (2) *takes place for every trajectory, there is a limiting cycle.*

For let M_0 be a first point, the consequents of which give a set P; N, a point of P'; Q, the set of the consequents and antecedents of N. If Q' coincided with P', (1) would be verified for the initial point N, which is contrary to the hypothesis. But, if Q' is distinct from P', we know that we must have a limiting cycle.

[1]The two cases are not distinct: if any trajectory is closed, every other trajectory must either be also closed or tend to a limiting cycle, this being seen in the same way as in the theory of centers (see p. 10; the argument is clearer if we suppose that the closure takes place after one revolution for φ and one revolution for ω, that is, $j=k=1$, but any other case can be reduced to that one, as has been remarked in footnote (1) on p. 42).

More precisely, we see that (1) is verified (in the absence of limiting cycles) for every initial point which belongs to the derived set of the consequents of another point included in P'. It is easily seen that P' is a perfect set.[1] We proceed to investigate the different kinds of perfect sets that may arise. There is first the possibility that P' contains an arc AB of the circle. We shall consider two cases here: (a) P' includes the whole circle; and (b) P' includes an arc AB, but not the whole circle.

We consider (b) first, and suppose AB cannot be enlarged without enclosing points not belonging to P'; that is, we will not consider a smaller arc when a larger one containing it may be chosen. Consider the consequent arc A_1B_1. If A_1B_1 is enclosed in AB, then A_2B_2, the next consequent, is enclosed in A_1B_1, and so on. The limit A' of the points $A, A_1, A_2, \cdots$, gives us a closed trajectory; so this case must be rejected.

If the arc A_1B_1 contains points of AB and also points outside AB then AB is included in a larger arc, which we supposed was not the case. Thus if the points appear in the order AA_1BB_1, the larger arc AB_1 is contained in P'. So we put this case aside.

There remain the case that A_1B_1 is outside AB. Then all consequents of AB are exterior to one another. It follows that all consequents of a point of P lying in AB are outside AB; hence no interior point of AB is a point of P'. This is contrary to hypothesis.

In all cases, then, we have proved that (b) is impossible. Hence if P' contains any arc, it contains the whole circle. Further if (a) holds for the set of consequents of a point

[1]We already know that P' is closed. On the other hand, if P is included in P', any point of P' is near an infinity of points of P, which are points of P'; and, therefore, it belongs to P''.

M, it holds for the consequents of any other point N of the circle (if limiting cycles are excluded), as has been shown above.

If we are in this case, we can give a precise answer to one of the questions set in the above. We have seen that, starting from the initial point M_0 the argument of which is ω_0, the various consequents of M_0 or, what comes to the same thing, the reduced values

$$\omega'_i = \omega_i - E(\omega_i - \omega_0) = \omega_i - E(\mu i) \tag{6}$$

occur in the same order as the corresponding values of

$$h_i = \mu i - E(\mu i) \tag{7}$$

the subtracted term being the same in both formulae. Now, let ω' be any given argument between 0 and 1. If it coincides with an ω'_i, we shall associate with it the corresponding value of (7). If not, let us notice that there will be an infinity of ω'_i less than ω' and an infinity of ω'_j greater than ω', every ω'_j of the second class being therefore greater than any ω'_i of the first. The same inequalities holding, as we have seen, between the corresponding values of (7), these (which, as is well known, are everywhere dense between 0 and 1) will be divided into two classes such that a Dedekind cut is obtained: by this cut a number $h = h(\omega')$ is defined.

The function $h(\omega)$ is defined over the interval (0,1), and continuous along this interval (for we can choose two values h_i and h_j of (7) including h and as near as we like to h; and, then, the function considered will be contained between h_i and h_j for every ω contained between ω'_i and ω'_j). It is never decreasing, in fact, under our present assumption that P is everywhere dense over the circle, it is constantly increasing. Since we can find values of (7) less than 1 and arbitrarily near 1, it tends to 1 when ω tends to

ω_0+1, so that, if we now extend its definition to every real value of ω by subjecting it to the identity

$$h(\omega+1)=h(\omega)+1, \tag{8}$$

its continuity will be maintained throughout $(-\infty,+\infty)$. On account of (8), the difference $\theta(\omega)=h(\omega)-\omega$ will be periodic in ω (so that it admits of a continuous Fourier expansion).

As (6), (7) and (8) give us

$$h(\omega_i)=\mu i,$$

we have, for every ω_i, and, therefore (by continuity), for every ω,

$$\psi(\omega)+\theta[\psi(\omega)]=\omega+\theta(\omega)+\mu. \tag{9}$$

By this equation—which, under our present hypotheses, can be solved for $\psi(\omega)$—the construction of the function $\theta(\omega)$ would allow us to solve every question relating to the law of consequence.

Such are the results when the perfect set P is everywhere dense. But we know (and, as we have said, Poincaré was the first to discover, though the fact was stated explicitly only by Cantor) that there are other perfect sets, which are nowhere dense. The question arises whether this might be the case for our set P', the derivative of the set P of consequents of M_0.

It is known that any perfect discontinuous set is obtained by the removal of the interior points of an infinite number of intervals ("contiguous intervals"), (everywhere dense on the circle) the extremities of which, however, belong to the set: the latter is constituted by these extremities and their limiting points. Thus, under our third hypothesis, the circle will contain three kinds of points:

1° Interior points of the contiguous arcs: these do not belong to P';

2° Extremities of contiguous arcs;

3° Limiting points of points 2°.

It is clear that the consequent of any contiguous arc is again a contiguous arc. We have seen that the set P' is the same for any choice of the initial point M_0. But according to that choice, three kinds of trajectories are possible, two of which are such that the corresponding P is included in P', while under the third one—that is, when the initial point is taken within a contiguous arc—P has no common point with P'. Such a distinction could not occur in the two preceding cases in which P was entirely exterior to P', whatever M_0 might be, or entirely included in P', whatever M_0 might be: so that the question of its possibility is the same as the question whether the perfect set P' can be discontinuous.

These two questions, equivalent to each other, were not solved by Poincaré. He did not prove the impossibility of P' being perfect and discontinuous, though he pointed out some circumstances in which this impossibility would be certain.

This fundamental problem has been recently solved—and thus Poincaré's theory brought to completion—in a quite simple and elegant manner, by a note of Denjoy in the *Comptes Rendus de l'Académie des Sciences*.[1] The third hypothesis, viz., that P' is perfect and discontinuous, is certainly excluded if a certain integral (which expresses the relative change in the distance between two neighboring trajectories when φ increases by 2π) considered as a function of the initial position M, is of bounded variation; and this is always the case when the coefficients of the equation are holomorphic—or even when they admit a sufficient number of derivatives.

[1]Vol. CXCIV, 830, March 7, 1932.

V

HIGHER DIFFERENTIAL EQUATIONS: SINGULAR POINTS AND THE THEOREM OF KRONECKER

IN the fourth part of Poincaré's memoir,[1] he considers differential equations of the second order. The equations can be put in the form

$$(1) \qquad \frac{dx}{dt} = X(x, y, z), \ \frac{dy}{dt} = Y(x, y, z), \ \frac{dz}{dt} = Z(x, y, z).$$

The trajectories are curves in space.

There arises again the question of singular points and of their classification according to the behavior of the trajectories in their neighborhoods. There is also the question of their existence, and I wish to state that for a sphere, for instance, there must be singular points. This result is, again, a matter of Analysis Situs. In the ordinary case, the question is connected with the famous theorem of Kronecker,[2] which we have already seen[3] to have been used by Poincaré, and which has proved such a magnificent tool in his hands.

If the functions X, Y, Z are (as we shall assume) everywhere regular—so that the system behaves like a "system of the second order and the first degree"[4]—the only singular points correspond to the solutions of the system of ordinary equations

$$(2) \qquad X = Y = Z = 0.$$

[1] *Jour. de Math.*, ser. 4, II, 151–217.

[2] *Monatsberichte of the Berlin Academy*, March and August, 1869.

[3] See my Lectures of 1920, p. 151, 163.

[4] Poincaré even assumes X, Y, Z to be polynomials.

Every *simple* solution of this system (that is, every solution for which the Jacobian does not vanish) can be treated in the same way as the simple singular points of the equations of the first order; but there will be three characteristic numbers s (instead of two) so that there may be not only nodes, passes, and foci, but also singular points of a mixed character, *pass-foci*.

Now, let S be any closed surface, which will be the boundary of a determinate volume V. Let us assume that no solution of the system (2) lies on S itself. Kronecker's theorem gives us information concerning the solutions lying within V. If we call such a solution a *positive* or a *negative* one according to the sign of the Jacobian, the theorem gives the value of the "index," that is, of $P-N$, P being the number of positive and N the number of negative solutions inside V. This number can be obtained by considering the distribution of the vector (X,Y,Z) at the different points of S (where, on account of the hypothesis, this vector never vanishes). We construct a parallel vector (with the same sense) from the center of a unit sphere, and in this way, determine a point on the sphere. Thus, any point m of the sphere is covered a certain number of times. The required index is given by the number of positive coverings diminished by the number of negative ones—a difference which, as is easily seen, is the same whatever the choice of m may be.

The results of the application of this theorem will be analogous to those met with for the first order,[1] but a little less satisfactory, as will always happen when the number of dimensions is odd (in other terms, when the order of the differential system is even), because changing the signs of all the characteristic numbers s will change the sign of

[1]Lectures of 1920, p. 177.

the Jacobian (which is their product), so that the sign of a singular point, according to the above definition, has no fixed relation with its nature.[1] But this is no objection to the most important use of the theorem, which is to prove the existence of at least one singular point.

The difficulty may lie in the study of the vectorial distribution (X,Y,Z) in order to calculate the index. From this point of view, Poincaré has considerably increased the bearing and usefulness of Kronecker's result by an important complementary theorem which has since been rediscovered, independently (in fact, in an apparently different form and with a different proof) in a beautiful work of Bohl. Consider two vector distributions (X,Y,Z) and (X',Y',Z'). Poincaré and Bohl prove that *either they have the same index or there exists a point on the surface such that*

$$\frac{X}{X'} = \frac{Y}{Y'} = \frac{Z}{Z'} < 0.$$

Poincaré proves this[2] by considering the vector

$$(\lambda X+\mu X',\ \lambda Y+\mu Y',\ \lambda Z+\mu Z'), \quad \lambda, \mu > 0.$$

If there is a certain ratio of λ and μ for which this vector vanishes, we have the equations and inequality just given. If not, the index of Kronecker does not change as we vary λ and μ in such a manner as to pass from the first distribution to the second.

Especially, if the surface is contactless,[3] this complementary theorem can be applied in taking for (X',Y',Z') the direction cosines of the outer normal (the outer normal or the inner only according to cases). For the direction

[1]The nature of a singular point depends on the *compared* signs of the numbers s.

[2]Bohl's proof is less simple.

[3]See Lectures of 1920, p. 182.

of the outer normal, Poincaré finds that the index is equal to the quantity

$$\frac{\text{Total Gaussian Curvature}}{4\pi}.$$

Of course also, such a number cannot change when the closed surface is deformed continuously, so that it must only depend on topological features. Poincaré finds it for the genera 0 and 1; generally, Walther Dyck has found that the total curvature K of a closed (everywhere regular) surface of genus p is

$$K = 4\pi(1-p).$$

VI

THE INTRODUCTION OF PERIODIC SOLUTIONS

THE continuation of the fourth part of the Memoir *Sur les courbes définies par une équation différentielle* acquaints us with two capital new ideas, the importance of which, in the development of Poincaré's thought, has become greater and greater in later years. The first one is the use of periodic solutions. The first scientist who pointed out the importance of this kind of solution was the astronomer Hill, in his work on the theory of the Moon. But Poincaré makes them a capital tool for further investigations. We shall see that by their help, local methods, such as those obtained by Newton, Cauchy, and those found in the first parts of the Memoir we are dealing with, can be used over an extended domain, namely, the neighborhood of the closed trajectory.

Suppose we know one solution of

$$(1) \qquad \frac{dx}{dt} = X, \quad \frac{dy}{dt} = Y, \quad \frac{dz}{dt} = Z,$$

which is a closed curve L. There will be solutions infinitely near this one, and their study will depend, as is now well known, on methods analogous, to a certain extent, to those used in the neighborhood of a single point. Let us, in the given system, replace x,y,z by $x+\xi$, $y+\eta$, $z+\zeta$, where ξ, η, ζ are small: we shall have

$$\frac{d\xi}{dt} = X(x+\xi, y+\eta, z+\zeta) - X(x, y, z) = \frac{\partial X}{\partial x}\xi + \frac{\partial X}{\partial y}\eta + \frac{\partial X}{\partial z}\zeta,$$

if we neglect terms of higher order; and similarly for $d\eta/dt$ and $d\zeta/dt$.

These three equations are Darboux's[1] "auxiliary system" or, as Poincaré more logically calls them, the *variational equations*. A fundamental fact is that these variational equations, which define the infinitesimal increments ξ, η, ζ as functions of t, are linear ones. Moreover, in the present case, the coefficients $\partial X/\partial x, \cdots$, are known periodic functions of t. Now, another fundamental circumstance is that such linear differential equation with periodic coefficients admit of what could be called "semi-periodic" solutions, that is, solutions of such a form as

$$\xi = \Xi e^{\alpha t}, \quad \eta = \mathrm{H} e^{\alpha t}, \quad \zeta = \mathrm{Z} e^{\alpha t},$$

where Ξ, H, Z are again periodic in t, and α is a constant. Indeed, with the exception of peculiar cases (such as occur in the quite similar theory of linear equations with constant coefficients, when the characteristic equation admits of equal roots), such solutions will be in sufficient number to allow the construction of the general integral, so that, in the above case, there will be three, with three (generally different) exponents, which will be called the "characteristic exponents."[2]

Here, however, the given system is really of order 2 (the variable t being, so to speak, superfluous) and we can manage to put this in evidence. Let t be considered as a parameter corresponding to each point of the closed given

[1] *Comptes Rendus*, XCVI, 766, March 19, 1883; see *Leçons sur la Théorie des Surfaces*, Vol. IV, note XI, p. 505.

[2] The values of these characteristic exponents are not known in advance, and their construction in terms of the given coefficients is a by no means insignificant difficulty of the question. In the classic case of Celestial Mechanics, however, the given equations being very slightly different from integrable ones, this determination can be done by expansions in power series with respect to the perturbating masses: see a very good Memoir of F. R. Moulton and Macmillan in the *Am. Jour.*, XXXIII (1910), 63.

curve (and, as is evidently possible, be assumed to vary from 0 to 2π by one description of that curve). At O, where $t=0$, let us construct a small piece of plane normal to the curve—a "target," or, in Birkhoff's terminology, a (local) "surface of section." At each point, M, of the closed curve, we construct a similar target. Each point in the neighborhood of the closed trajectory lies on a unique target. We choose axes in the target with the origin at O. A point in the neighborhood of the closed trajectory will then be represented by the coordinates t,x,y, where the t is that of the target and x and y are the coordinates of the point in the plane of the target.

We define a trajectory near the closed trajectory by assigning a starting point m on the target at O and lying near O. Here x and y will be small and the trajectory will follow rather closely the closed trajectory. The new trajectory will meet the various targets as the motion proceeds, and will return to the original target at O when $t=2\pi$. Let us direct our special attention to this second point of impact $m_1(x_1,y_1)$. We shall speak of m_1 as the first consequent of m. We shall have a point-transformation

$$x_1=\Phi(x, y), \quad y_1=\Psi(x, y),$$

where Φ and Ψ vanish at the origin and, if analytic or at least sufficiently regular, can be written in the form:

$$x_1=ax+by+\cdots, \quad y_1=cx+dy+\cdots. \tag{2}$$

The trajectory will meet again the initial target at a third point m_2, the consequent or m_1, or second consequent of m; and so on. The problem is therefore reduced to the investigation of the distribution of these successive consequents, in other words, to the iteration of our point transformation. We shall speak not only of Poincaré's work in this line, but also of the most important improvements

which have been given to it, especially in Birkhoff's fundamental paper *Surface Transformations and their Dynamical Applications*,[1] which is both rich in results and suggestive of further researches. As a first approximation, we consider only the linear terms. These can be reduced, by a suitable change of variables, to the canonical form

$$x_1 = s_1 x, \quad y_1 = s_2 y, \tag{3}$$

where s_1 and s_2 are the roots of

$$\begin{vmatrix} a-s & b \\ c & d-s \end{vmatrix} = 0,$$

provided these roots are real and distinct.

In the exceptional case in which the roots are equal we can no longer, at least not in general, put the linear part of our transformation in the form (3). But we can then obtain the form

$$x_1 = sx, \quad y_1 = s(x+y). \tag{3a}$$

If the roots are conjugate imaginaries, so are the variables (3). But, by a proper choice of real variables and the introduction of polar coordinates, we can obtain

$$\rho_1 = s\rho, \quad \omega_1 = \omega + \alpha, \tag{3b}$$

where s is positive. This is a rotation combined (at least if $s \neq 1$) with a homothecy, s being the common value of the moduli of s_1 and s_2.

Of course, a method to calculate these linear terms in (2) would be to integrate the variational equations. Especially, in (3), s_1 and s_2 will be connected with the above defined characteristic exponents α_1 and α_2 by the relations

$$s_1 = e^{2\pi\alpha_1}, \quad s_2 = e^{2\pi\alpha_2}.$$

Returning to equations (3), we see that the effect of the iteration will depend upon the magnitude of s_1 and s_2.

[1] *Acta Math.*, XLIII (1920).

(I) If the absolute values of both are less than 1, the successive consequents will be nearer and nearer the origin. Moreover, if $|s_2| < |s_1|$, y_n decreases more rapidly than x_n and the successive consequents lie on a curve tangent to the x-axis at the origin; that is, unless $x=0$ initially, in which case all the consequents lie on the y-axis.

It is easy enough to see that every essential part of this remains valid if we take account of higher terms. The successive consequents of a point will tend to the origin, and, usually, in a direction tangent to the x-axis.

The last statement will fail if the initial point lies on a certain line tangent to Oy (no longer the y-axis exactly), which is invariant by the transformation.

That such an invariant line actually exists, results, in the analytic case, from a proof given by Poincaré for the case in which $s_2 < 1 < s_1$.[1] Let us write the (complete) equations of the transformation in the form

$$(\mathcal{T}) \qquad x_1 = s_1 x - \Phi_1(x, y), \quad y_1 = s_2 y + \Phi_2(x, y),$$

Φ_1, Φ_2 being holomorphic expansions in powers of x,y, the terms of least order in which are quadratic. A curve (assumed to be tangent to Oy)

$$(4) \qquad x = g(y) = \alpha_2 y^2 + \cdots + \alpha_m y^m + \cdots,$$

becomes, by this transformation, $s_1 x - \Phi_1 = g(s_2 y + \Phi_2)$, so that it will be invariant if we have

$$g(s_2 y + \Phi_2) = s_1 g(y) - \Phi_1.$$

We write this equation so as to solve it for the combination

$$(5) \qquad s_1 g(y) - g(s_2 y),$$

namely,

$$(5a) \qquad s_1 g(y) - g(s_2 y) = g(s_2 y + \Phi_2) - g(s_2 y) + \Phi_1.$$

[1] See page 66 of this lecture.

Then, if we introduce the unknown coefficients of the Maclaurin expansion (4) of $g(y)$, we see that each of them will appear in the right-hand side in higher terms than in the left-hand one, so that we can, in the usual way, consider (5a) as giving the quantity (5) just as if the right-hand side were completely known: each coefficient α_m will be given by an equation of the first degree in which its coefficient will be the corresponding value of

$$s_1 - s_2^m \qquad (m = 2, 3, \cdots, \infty). \tag{6}$$

Moreover, these successive calculations will introduce only additions and multiplications, without any negative terms (the subtractive part in the right-hand side of (5a) having no other effect than to cancel some parts of the first term): therefore, the method of dominating functions can be applied. As Φ_1 and Φ_2 are assumed to be holomorphic, and of order two in x,y, we can admit that

$$\Phi_i < < \frac{M\beta^2(x+y)^2}{1-\beta(x+y)} = \Phi(x, y), \qquad (i = 1, 2), \tag{7}$$

with proper values of M and β, and convergence will certainly take place for the given problem if it does when we substitute for Φ_1 and Φ_2 these dominating values (7), simultaneously replacing s_1—and s_2 if negative—by $|s_2|$ (which has the effect of diminishing each of the numbers (6) and, consequently, of increasing each α_m). But this is equivalent to finding an invariant line for the transformation

$$\begin{aligned} x_1 &= |s_2| x - \Phi(x+y), \\ y_1 &= |s_2| y + \Phi(x+y). \end{aligned}$$

Now, the latter gives $x_1 + y_1 = s_2(x+y)$, so that if we write the equation of our line in the form

$$x = \psi(x+y), \tag{8}$$

the condition of invariance will be given by the functional equation in ψ

$$\psi(|s_2|u) = |s_2|\psi(u) - \Phi(u)$$
$$= |s_2|\psi(u) - M(\beta^2 u^2 + \beta^3 u^3 + \cdots + \beta^m u^m + \cdots).$$

This has a holomorphic solution such that $\psi(0) = \psi'(0) = 0$, namely,

$$\psi(u) = \frac{\beta^2 u^2}{|s_2| - |s_2|^2} + \frac{\beta^3 u^3}{|s_2| - |s_2|^3} + \cdots + \frac{\beta^m u^m}{|s_2| - |s_2|^m} + \cdots .$$

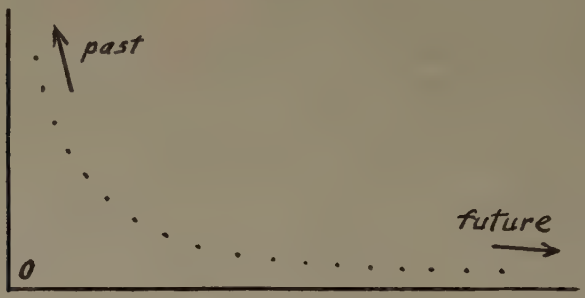
Since, for this choice of ψ, (8) can be solved for x, the existence and convergence of the expansion (4) is proved.

The interpretation of these results in space is simple. As the consequents approach O, the successive arcs of the trajectory come nearer and nearer to the periodic trajectory through O. The trajectories then approach the periodic solution asymptotically, all in the same direction—that is, tangentially to one and the same surface—except for some exceptional ones, which meet our target at points of (4) and, therefore, lie on another determinate surface through L.

(I′) If s_1 and s_2 are both greater than 1 in absolute value, the successive consequents recede from O and our present method tells us nothing about the trajectory as t increases, except that it departs from the neighborhood of L. But by reversing the sign of t and considering the antecedents of m, our treatment again applies. The trajectory is asymptotic to the periodic solution for decreasing t. Evidently, all that we have just said for case (I) applies to our present one, by changing t into $-t$.

(II) Similar remarks apply to the case (3b) of conjugate roots (except those which concern invariant lines). If $s < 1$, the trajectory is asymptotic to the closed trajectory for the future; if $s > 1$, it is asymptotic for the past. In this case, there is a spiral motion about the closed trajectory.

(III) Suppose now $s_1 > 1$ and $s_2 < 1$, the "hyperbolic case" in Birkhoff's language. Then, considering the approximate equations, the consequents have, in general, increasing abscissas and decreasing ordinates, whereas the antecedents have increasing ordinates and decreasing abscissas. The antecedents and consequents form a figure like the following on the target:



If, however, the abscissa of m is zero, the consequents approach O along the y-axis. Similarly, the antecedents of a point on the x-axis approach O.

Again, similar conclusions will hold for the exact equations in which we take account of higher terms. This has been shown by Poincaré himself and in several later works, especially a valuable thesis of S. Lattès[1] and the above mentioned Memoir of Birkhoff.

Poincaré himself assumes the transformation to be analytic, and it is for this case that he constructs an invariant curve tangent to Oy by the method which we have explained above and which, as we have seen, is also valid for the case (I). In case (I), that method gave one invariant curve; in our present one, it gives two, because we can treat in the same way the inverse transformation, whereby

[1]Paris, 1906. It has been published in the *Annali di Matematica.*

s_1, s_2 will be changed into $1/s_2$ and $1/s_1$. Thus, we shall have a second invariant line tangent at O to Ox.

In three-dimensional space, this gives two surfaces through L, one of which is the locus of trajectories asymptotic to L for positive t's, the other being the locus of trajectories asymptotic to L for negative t's.

The above considerations would incline us to believe that the second invariant line, tangent to Ox, does not exist in case (I). However, this would be an error. In his thesis, Lattès has proved, with the help of a proper device, that the only necessary assumption for its existence—instead of $s_1 > s_2$—is (s_2 being still less than 1) that none of the numbers (6) vanishes (the contrary case of $s_1 = s_2^m$ being analogous to that of equal roots and reducible to it).[1]

This is what we can obtain in the analytic case by the use of entire series. But I must say that, in my opinion, there is always an advantage in completing results attained in this way by direct proofs, based on methods of successive approximations. Not only do such proofs allow us to get rid of one hypothesis, namely, that of analyticity; but they are always more instructive, showing us, as we shall see in the present example, the behavior and mechanism of the phenomenon, which are not brought out when we use power series. Moreover, the question whether analyticity is implied or not in the solution of a problem has in itself a great importance: it penetrates profoundly into the nature of the problem, as I have been led to see by my own researches on partial differential equations.

Now, I was able to prove[2] that the existence of two invariant lines (and only two, in the present case of $s_1 > 1 > s_2$)

[1] Lattès has also extended the results to three-dimensional transformations.

[2] *Bull. Soc. Math., France, Comptes Rendus des Séances*, 1901, 324. As in many other questions concerning iteration, the method used there is not essentially different from an "on and back" method (see page 83 of this lecture).

is certain for any (sufficiently regular) transformation (T), whether analytic or not. Let us take *any* line L_0 through O not tangent to Oy, and its successive transforms by (T) iterated an infinite number of times. We shall see that these tend to a perfectly definite limiting position, *which is independent of the choice of the initial line* L_0, and tangent to Ox. In fact, we could start, not with a line L_0, but with a whole region—say a circle—surrounding O: if we construct the successive transforms of such a region, we can show that they will be contracted in one direction (roughly, that of Oy) and protracted in another one, so that they (or more exactly, their parts not too remote from O) will degenerate into a single line, which is the desired one. It appears, hereby, that this is an invariant curve and that it is the only one not tangent to Oy.

If, moreover, we assume that $s_1 > 1$, we can proceed in the same way in the opposite sense, the only difference being that we no longer iterate the transformation (T), but its inverse: in other words, we shall now consider the successive antecedents of the arbitrary line L_0, instead of its consequents. L_0 being assumed, this time, not to be tangent to Ox, these antecedents will approach another limiting curve, which will be tangent to Oy. We thus see that there are two and only two invariant curves.

When the point m describes one of these curves, the corresponding trajectory will describe one or the other of two sheets of a surface containing our closed curve L. Such trajectories will approach L asymptotically either for positive or for negative t's; and such is the case for no other trajectories.

In the contrary case where unity does not lie between s_1 and s_2, say $s_2 < s_1 < 1$, the above process affords one

invariant curve (instead of two), which is tangent to Oy: Lattès has clearly characterized its significance by observing that it is the only invariant curve through O which is not tangent to Ox, or also, the locus of a point P the consequents of which approach O otherwise than tangentially to Ox. As to invariant curves tangent to Ox, he remarks that there exists an infinite number of them: for, P_0 being any point in the neighborhood of O and $P_1, P_2, \cdots$, its successive consequents, we can connect P_0 to P_1 by an arbitrary arc, the consequents of which will connect P_1 to P_2, P_2 to P_3, etc., and, finally abutt at O, since the consequent P's have O as their limiting point. We can easily choose the initial arc P_0P_1, moreover, so that the total curve thus constituted possesses a continuous tangent, or even a continuous curvature, etc. In other words, the problem of the invariant line (with a tangent at O other than Oy) is not clearly set if we do not add proper conditions of regularity which, in the general case, would be very difficult to formulate, as soon as analyticity fails to intervene.

We may say, by the above considerations, that the main questions are solved whenever the coefficients s_1 and s_2 are different from unity in absolute value.

In all these cases, moreover, we see a striking analogy with what was found concerning singular points of equations of the first order and the first degree. In each of them, successive consequents may be considered as marked along one of the curves which we drew previously (see p. 10), the first case examined ($s_2 < s_1 < 1$) corresponding to what takes place around a node, the second one ($s_1 > 1 > s_2$) to a pass. The relation between the two theories is that of the continuous to the discontinuous: the two series of diagrams would become respectively identical to each

other[1] if, in the previous one (p. 10), we should consider only, for instance, points corresponding to integral values of the variable t.

Such analogies, which lie in the nature of things, will subsist, but with important differences—in fact, new and profound difficulties—in the remaining cases where the moduli of s_1 or s_2 are equal to 1, according as these numbers are real and equal to ± 1, or are conjugate imaginaries.

This is what Poincaré finds on attacking the latter case, an especially important one for comparison with problems of Celestial Mechanics. In this case, the linear part of the transformation is given by (3b), but with $s = 1$, that is

$$\rho_1 = \rho,$$
$$\omega_1 = \omega + \alpha.$$

For these approximate equations, the successive consequents lie on a circle with O as center. The trajectory would then remain on a certain tubular surface enclosing the closed trajectory.

The study of the equations with the terms of higher order included will, therefore, correspond to the study of centers, as above discussed for the equation of the first order. Let us see how our analogies and differences will behave.

We no longer start, this time, with the point transformation (T), but with the differential equations themselves. If, in the first place, we reduce them to the variational equations, they will have integrals of the semi-periodic form written above and, in the case which we now examine,

[1]In the system of coordinates just used in the text, the differential equations, giving x and y (coordinates on the successive targets) in terms of t, are exactly of the same form as the equations considered for the first order around a singular point, except that the right hand sides X, Y are no longer independent of t, but periodic functions of it. That our present case would give similar results to those formerly found, but with greater generality, was therefore to be foreseen.

the characteristic exponents are conjugate pure imaginaries, say

$$\alpha_1 = i\lambda, \quad \alpha_2 = -i\lambda,$$

so that the system will admit of two solutions of the form

$$A(t) \cos \lambda t + B(t) \sin \lambda t$$

(A,B, periodic functions with the period 2π), the combination of which gives the general integral, as usual.

We easily deduce therefrom a linear change of variables on x and y, its coefficients being periodic in t with the period 2π, such that it reduces the variational equations to

$$\frac{dx}{dt} = \lambda y, \quad \frac{dy}{dt} = -\lambda x.$$

By such a choice of variables (obviously a generalization of the classic method of "variation of constants"), the complete equations (where account is taken of higher terms) acquire the form

$$\frac{dx}{dt} = X = \lambda y + X_2 + \cdots + X_n + \cdots,$$
$$\frac{dy}{dt} = Y = -\lambda x + Y_2 + \cdots + Y_n + \cdots,$$

as written in the theory of centers (see page 9), except that it contains t explicitly, with respect to which X and Y are periodic with the period 2π (and that the first terms contain the numerical factor λ).

Imitating our previous method, we shall try to find an integral

$$F = F_2 + F_3 + \cdots,$$

that is, a function satisfying the partial differential equation

$$X\frac{\partial F}{\partial x} + Y\frac{\partial F}{\partial y} + \frac{\partial F}{\partial t} = 0.$$

If we can find a function F, where (in polar coordinates ρ and ω)

$$(9) \qquad F_q = \rho^q \Phi_q = \rho^q \Sigma[a \cos(mt+n\omega) + b \sin(mt+n\omega)],$$

where Φ_q is a trigonometric series in t and ω, m and n assuming the various integral values, the general equation $F = \text{const.}$ will define a family of tubular surfaces enclosing our given line L, each of which will be a locus of trajectories.

Now, again, we easily find $F_2 = \rho^2$ and if we assume that the first $q-1$ terms have been found in the form (9), the q^{th} one will have to satisfy an equation

$$(10) \qquad \frac{\partial \Phi_q}{\partial t} - \lambda \frac{\partial \Phi_q}{\partial \omega} = H_q,$$

where the expression H_q in the right-hand member is analogous to (9), say

$$H_q = \Sigma[A_{m,n,q} \cos(mt+n\omega) + B_{m,n,q} \sin(mt+n\omega)].$$

Just as in our theory of centers, we have two kinds of questions before us:

1° whether each successive equation (10) can be solved by an expression of the form (9);

2° if so, whether the terms thus found will be those of a convergent series.

For the first question, the same conclusions as in our previous problem will hold. An impossibility will arise whenever one of the expansions H contains a constant term C_0 different from zero. If so, we shall be able to construct a family of tubular contactless surfaces, and instability will be certain, each trajectory being made to approach L asymptotically and spirally either for positive or for negative t's.

Again the question is, therefore, whether every C_0 vanishes. But this time, the question may be regarded as

solved *a priori*, at least for all important cases from the point of view of Analytical and Celestial Mechanics. This is the consequence of the second capital concept introduced by Poincaré in this fourth Memoir: the notion of integral invariants.

As has happened in some of the most important scientific discoveries, the initial principle of it was found independently by several men. Boltzmann and Gibbs introduced it in the kinetic theory of gases, and it is now known that its true author is Liouville. It can be proved at once with the help of the modern Vector Theory. Let us resume the former meaning of x,y,z as coordinates in three-dimensional space. Consider our differential system

$$\frac{dx}{X}=\frac{dy}{Y}=\frac{dz}{Z}$$

as defining a vector field and look upon the integral curves as lines of flow. If the divergence is zero, the motion is that of an incompressible fluid. If we draw the trajectories from the points of a volume v from a time t_0 to a time t_0+T, the terminal points will fill a volume v_1. The two volumes will be equal in magnitude if the divergence is zero; that is, if

$$\frac{\partial X}{\partial x}+\frac{\partial Y}{\partial y}+\frac{\partial Z}{\partial z}=0.$$

More generally if

$$\frac{\partial(MX)}{\partial x}+\frac{\partial(MY)}{\partial y}+\frac{\partial(MZ)}{\partial z}=0,$$

where $M=M(x,y,z)>0$ say, then $\iiint M dxdydz$ is an integral invariant. Now, if there is an integral invariant, there can be no contactless closed surface. Thus an incompressible fluid could not flow into a closed surface at

all points of the surface, as this would imply an augmentation or diminution of its volume. The contrary case, which is physically impossible, could only occur analytically if singular points existed inside the closed surface (what hydrodynamicists call "sources" or "sinks").

This solves our question concerning the C_0's, if such an integral invariant exists: then, the neighborhood of L being assumed to be devoid of singular points, no closed contactless surface can be constructed about L, so that we are assured *a priori* that every C_0 is zero.[1]

Such is, precisely, the case in all problems which are likely to interest us from the dynamical point of view: for it can be seen at once that Hamilton's equations always admit of an integral invariant, the volume itself, the above quantity M being equal to 1.

Thus, we know in advance that the first kind of difficulty will never stop us, and this is in itself important in relation to the modern development of Celestial Mechanics. For the above mode of reasoning is exactly parallel to a celebrated method due to Lindstedt. Now, in the latter method, it was a question of showing that "secular terms" would never appear; that is what Lindstedt did, but with some restrictive hypotheses. Poincaré's argument proves at once this property of Lindstedt's method, even without the restrictions introduced by his predecessor.

The possibility of constant terms in the H's being excluded, each of the F's in (9) will be a sum of terms respectively defined by such equations as

[1]The same deduction would have been possible in the theory of centers; but it would have been useless because, for an equation of the first order, the knowledge of the integral invariant is equivalent to that of the classic "integrating factor" and, therefore, to the direct integration of the equation. (In the general case, the factor M in an integral invariant is nothing else than Jacobi's "multiplier.")

$$\frac{\partial \Phi}{\partial t} - \lambda \frac{\partial \Phi}{\partial \omega} = A \cos(mt + n\omega + \theta), \tag{10'}$$

A, θ being given constants.

Now, one solution of such an equation is, in general,

$$\Phi = a \sin(mt + n\omega + \theta), \tag{9'}$$

with

$$a = \frac{A}{m - \lambda n}. \tag{9a}$$

If λ *is incommensurable*, this will always be possible, and our calculations can be pursued indefinitely without meeting with any difficulty.

On the contrary, for a commensurable λ, the denominator $m - \lambda n$ will vanish an infinite number of times, when we give m and n all possible integral values. If the corresponding right-hand side of (10′) is not zero, (10′) admits of no trigonometric solution, and a new kind of secular term will appear, containing t multiplied by sines or cosines. This is the well known difficulty of commensurable mean motions in Celestial Mechanics.

But the capital difference with our previous problem of centers will appear in our second question, namely, convergence. Contrary to what we have found concerning centers, we shall see that it must be, this time, answered negatively. This is due to the fact that λ, even if not commensurable, is always at least approximately so, that is, admits of an infinite number of approximate expressions by ordinary fractions. This is the difficulty of "small divisors," also a well known one in Celestial Mechanics.

The question is whether this difficulty is due to our methods of calculation, or, on the contrary, lies intrinsically in the nature of things. To attack this question we

shall, of course, take some relatively simple instances: it will be sufficient to consider the system, (ρ,ω still meaning polar coordinates on our targets)[1]

$$\frac{d\omega}{dt} = -\lambda, \quad \frac{1}{\rho(1-\rho)}\frac{d\rho}{dt} = \frac{d}{dt}\left(L\frac{\rho}{1-\rho}\right) = R = \Sigma A_{m,n}\cos(mt+n\omega),$$

the trigonometric series being assumed, for instance, to be uniformly convergent, which will certainly take place if $\Sigma A_{m,n}$ converges absolutely. The choice of the function $L\rho/(1-\rho)$ is such that it becomes equal to $+\infty$ at O and to $-\infty$ for $\rho=1$. Moreover, such a system is immediately integrable by quadratures, the first equation giving

$$\omega = -\lambda t + \omega_0,$$

by means of which the second one gives

$$(11) \quad L\rho/(1-\rho) = \int R dt, \quad R = \Sigma A_{m,n}\cos[(m-\lambda n)t - n\omega_0].$$

Now, several cases may be of interest for us:

1° We mention, to begin with, the case where there is a constant term $A_{0,0}$ different from zero. Let us admit that λ is incommensurable. Then, as previously, Poincaré shows that there is necessarily instability by constructing, around L, a family of tubular contactless surfaces. According to the sign of $A_{0,0}$, these surfaces

$$\mathrm{F} = \mathrm{c} = \text{const.}$$

—the above equation being written in such a way that $F-c>0$ outside the surface and $F-c<0$ inside—will be *positive* (which means that the total derivative dF/dt will be >0), or they will be *negative*. In the first case, every

[1]Poincaré's form is slightly different: the lines which he denotes by $\rho=$const. are also circles; but, instead of being concentric, they have a common radical axis, A, L being assumed to be itself a circle generated by the revolution of O around A.

trajectory will cross these surfaces outwards, and, finally, ρ will tend to unity; in the second one, every trajectory will cross them inwards and ρ will tend to zero.

2° Next, let us admit that $A_{0,0}$ is zero, but $\lambda = p/q$ is commensurable. If so, though R contains no constant term, R will contain, as a rule, an infinity of them. Let

$$K = \sum_{m/n=\lambda} A_{m,n} \cos n\omega_0$$

be their sum. In every other term of (11), the coefficient $m - \lambda n$ will be an integer, so that the integration of this term again gives an absolutely and uniformly convergent trigonometric series. But K will, as a rule, be different from zero, and will give a secular term. Therefore, ρ will tend either to 1 or to 0; but whether one or the other will depend, this time, on the sign of K, that is, on the value of the initial constant ω_0. There is again instability, except for some exceptional trajectories (those which correspond to $K = 0$ and which are even closed); but it is clear that no family of contactless surfaces can exist, as their presence would be contradictory to the last fact here noticed.

3° Now ($A_{0,0}$ being still zero), let λ be incommensurable. On account of the small divisors met with whenever m/n is an approximate value of λ, the series

$$\sum A_{m,n} \frac{\sin[(m - \lambda n)t - n\omega_0]}{m - \lambda n} \tag{12}$$

may be divergent and, thus, of no use for the study of $\int R dt$.

But another circumstance may occur: it may happen that the above series is convergent, but not uniformly; and this brings us back to the results previously obtained

by Poincaré in one of his first Memoirs in the *Bulletin Astronomique*.[1] Especially, it follows from that Memoir that, if the coefficients

$$\frac{A_{m,n}}{m-\lambda n} \tag{12'}$$

of the series (12) are not bounded, this series, even if always convergent, cannot remain finite for indefinitely increasing t's. Now, it is easy to see that such circumstances can actually occur. Let us take $\omega_0=0$. Then, if the series constituted by the coefficients $A_{m,n}$ in the right-hand side of (11) converges absolutely, so will (12) because $|\sin x/x|<1$. On the other hand, the order of magnitude of $m-\lambda n$ will depend on the closeness of the approximation of λ by the rational numbers m/n, which, as is well known, can be always disposed of arbitrarily by properly choosing the incomplete quotients of the expansion of λ in a continued fraction.

Poincaré also treats the question whether ρ will increase indefinitely or whether alternations of very great and finite values may occur: a further and more difficult discussion shows him that both cases are actually possible. But the important fact is that, in spite of the vanishing of all the C_0's and the possibility of performing the whole calculation of the series (12), even if the series is convergent, we can have instability, the new trajectories tending on the one hand, to recede more and more from L or, on the other, to approach it asymptotically on account of the presence of small divisors. Thus Poincaré is justified in finishing his Memoir with the following paragraph:

"D'après ce qui précède, on comprendra sans peine à quel point les difficultés que l'on rencontre en Mécanique Céleste

[1]See my Lectures of 1920, p. 159–162.

par suite des petits diviseurs et de la quasi-commensurabilité des moyens movements, tiennent à la nature même des choses et ne peuvent être tournées. Il est extrêmement probable qu'on les retrouvera, quelle que soit la méthode que l'on emploie."

VII

WORK ON POINT TRANSFORMATIONS AFTER POINCARÉ

THE paradoxical circumstances thus met with as to our differential equations must, of course, correspond to analogous facts concerning point transformations, for we know that the two topics are closely connected.

This is, nowadays, especially clear when we consider point transformations in their whole domains of definition. Indeed, the subject was proposed by the Academy of Sciences of Paris as a prize question for the year 1918, which gave rise to the works of three young geometers (two of whom we have since lost prematurely): Fatou, Lattès, and Julia. It was found that in the case which may be considered the simplest of all—transformations of one complex variable—transformations as simple as

$$z_1 = 2z^k + 1 (k \text{ an integer}),\quad z_1 = \frac{z+z^2}{2},\quad z_1 = \frac{3z-z^3}{2},$$

the most subtle and abstruse considerations of the modern theory of sets are necessary in order to define the regions in which the initial point must be taken in order that its successive consequents approach some one invariant point rather than another. The boundaries of such regions can by no means be considered as "lines," from the older and usual point of view: their study no longer depends on "combinatory Analysis Situs," but on "Analysis Situs of point sets."[1]

[1]See p. 37.

We can foresee from these facts that something of this complexity must penetrate even into the local domain surrounding an invariant point; and this appears clearly in some later Chapters of Poincaré himself[1] as well as in important works of other geometers.

We have seen that the most difficult and obscure case occurs when the moduli of s_1 and s_2 are no longer different from unity. In fact there is one subcase where these difficulties are most easily foreseen: namely, when our transformation is of the form

$$x_1 = x + \cdots, \quad y_1 = y + \cdots,$$

(dots standing, as usual, for terms of higher order), in other words, when its linear part reduces to the identical transformation, so that the distance between any point near the origin and its transform is an infinitesimal of the second order. It is evident that the cumulative and final effect of iterating such an operation will be more difficult to ascertain than in any other case. A fundamental question is the following: will this transformation be stable or unstable? That is, a point P being taken in the neighborhood of O, will its successive consequents remain also in the neighborhood of O or will they not?

It is clear that no answer to this question follows from what we have said thus far. But an answer has been given in two short Notes of Levi-Cività in the *Comptes Rendus de l'Académie des Sciences* of Paris,[2] in which the proof is afforded that, apart from exceptional cases, the transformation is *unstable*. The result may depend, as often happens, on the special choice of P in the vicinity of O; but we can indicate a certain angle A at O such that, if P lies within A,

[1] *Les Méthodes nouvelles de la Mécanique Céleste*, Vol. III, Ch. XXIII.

[2] July 9 and 16, 1900.

so will all its consequents, as long as they have not emerged from a certain corresponding circular sector S, from which they must necessarily emerge, and that only by crossing its circumference.

From this statement on point transformations, we can, of course, deduce a corresponding one concerning certain classes of periodic solutions.

But the most important improvements upon Poincaré's work, in that line as in several others, are due to Birkhoff, whose Memoir of the *Acta Mathematica* Vol. XLIII, cited above, is mainly devoted to this question of point transformations about a singular point, under the hypothesis that there exists an integral invariant, a restriction which, as we have seen, does not represent any diminution in the interesting character of the dynamical results.

The starting point of Birkhoff's method is the connection, pointed out by Poincaré, as we have seen above, between a point transformation and a continuous motion. We have noticed that the former can be deduced from the latter; but, conversely, when we are given a point transformation T, we can try to regard it as belonging to a continuous family, each point thus moving on a certain line between its original and its final position, say when the time t varies from 0 to 1, further integral values of t corresponding to the iteration of our given transformation if positive, and of its inverse if negative. Birkhoff does not investigate at all how this can actually be done in the most general way and constructs only one continuous set of such transformations, the equations of which can be explicitly written in terms of those of T. One of the most striking features of his argument—also connected, as we shall see below, with Poincaré's principles—is that the construction is purely

formal, that is, it gives series which are, as a rule, divergent and that, nevertheless, it enables him to obtain quite effective and rigorous results.

This is obtained by a kind of "on and back" method, which proves to be rather general. Indeed, it can be said to be the origin of practically every result in this kind of question. But it is Birkhoff himself who has put into full evidence the very nature and gist of it, especially when he applied it to such problems as equations in finite differences.[1] Let us suppose that T is unstable, so that, conversely, the iteration of T^{-1} will carry an arbitrary point P (at least if chosen within a proper region R) into the immediate neighborhood of the invariant point O. Let us suppose, moreover, that we know an approximation T' of T, that is, a transformation T' which differs from T only by terms of sufficiently high order in the neighborhood of O, and a line L' invariant under T' and located in R. Then, the transform P_k of P by the operation T^k will (if T' is a sufficiently close approximation to T) approach a limiting position p for $k=\infty$, and the locus of p, when P describes L', is a line invariant under T.

In the first place, he thus obtains a new solution for the hyperbolic case ($s_1>1>s_2$) previously treated (see p. 66) by Poincaré and Lattès. Choosing the variables so that T' reduces to[2] $x_1=s_1x, y_1=s_2y=y/s_1$ so that L is the x-axis, he finds that the series which give P_k admit a fixed dominating one which is convergent, and that they, therefore, admit of a limiting integral series.

So far, we have only a new and elegant solution of a problem already solved in other ways. But the results are

[1] See, for instance, his Memoir in the *Trans. Am. Math. Soc.*, XII (1911).

[2] In Birkhoff's cases, on account of the existence of the integral invariant, we always have $s_1s_2=1$.

of a quite new and remarkable nature when we attack Levi-Civitàʼs case

$$x_1 = x + \cdots, \quad y_1 = y + \cdots.$$

Again submitting the diagram to a preliminary transformation so that the approximate invariant line becomes the x-axis,[1] we find, if the order m of approximation (that is, the order of contact between $\mathcal{T}'$ and $\mathcal{T}$) is sufficiently great,[2] that the on-and-back method again gives the desired limiting invariant curves which are the only ones possible.[3] But, this time, nothing allows us to assert that this curve will be analytic at the invariant point: y may, and usually will, be a function of x belonging to the class which Birkhoff calls *hypercontinuous*, its expansion in an integral series being only an asymptotic one, with the radius of convergence zero. That such can be the case, is easily shown by the very simple example:

$$(\mathcal{T}) \qquad u_1 = \frac{u}{1+u}, \quad v_1 = (1+u)^2(v+u^2),$$

the direct iteration of which gives, for every integral k,

$$u_k = \frac{u}{1+ku},$$

$$v_k = (1+ku)^2\left\{v + u_2\left(1 + \frac{1}{(1+u)^4} + \cdots + \frac{1}{[1+(k-1)u]^4}\right)\right\}.$$

This can be written, with the help of the eulerian function Γ and its logarithmic derivative

[1] It proves more convenient, in this case, to do this so as to remove the invariant point to infinity in the x direction.

[2] The order m must generally be taken greater than 5, while, in the problem of Poincaré and Lattès, it was sufficient to take $m=2$.

[3] The number of such real curves may be either one or three, their tangents being determined by a cubic equation.

$$v_k = (1+ku)^2\left\{v+\frac{1}{6u^2}\left[\psi'''\left(\frac{1+ku}{u}\right)-\psi'''\left(\frac{1}{u}\right)\right]\right\},$$

$$\psi(z) = \frac{d}{dz}\log\Gamma(z),$$

since $\psi(z+1) = 1/z+\psi(z)$. It is then found that the existence of an analytic and regular invariant function or an analytic and regular invariant curve would be contradictory to the fact that the classic asymptotic series for the expansion of $\log\Gamma(1/u)$ (or any derivative of it) in powers of u is never convergent.

Poincaré foresaw in the *Méthodes nouvelles de la Mécanique Céleste* not only that such asymptotic expansions must intervene, but that their investigation might lead to effective and important results. We see that this is fully confirmed by Birkhoff's Memoir.

Further important and delicate researches are pursued by Birkhoff in the same Memoir, and are likely to shed a new light upon the integration of systems of the second order about a periodic solution.

The above considerations—not only the last part of Poincaré's Memoir but also the aforementioned continuations contributed by his successors—allow us to grasp the general importance of periodic solutions: an importance which has proved greater and greater on account of the results contained in such works as the prize Memoir of the *Acta Mathematica* and the *Méthodes nouvelles de la Mécanique Céleste*. Until a few years ago, these results even inclined us to believe that periodic solutions might generate all other ones—at least, those remaining in a finite domain—in the same way that rational numbers generate irrational ones by a limiting process. This, however, is not exact, as Birkhoff's works indicate: this rôle does not, in the general case, belong to periodic motions, but to more general

ones, the "recurrent motions" of Birkhoff. A striking fact, recently proved by Franklin, is the way in which this new and fundamental class of motions is connected with Bohr's almost periodic functions mentioned in the beginning (see p. 5).

But we must not forget that these new conceptions could hardly have arisen without having been prepared by the previous principles of Poincaré, whose sentence[1] remains true: Les solutions périodiques se sont montrées "la seule brèche par laquelle nous puissions essayer de pénétrer dans une place jusqu'ici réputée inabordable."

J. Hadamard.

[1] *Les Méthodes nouvelles de la Mécanique Céleste*, I, 82.

THREE PUBLIC LECTURES ON MODERN PHYSICS[1]

I

WHAT THINGS ARE MADE OF

TWENTY-FIVE HUNDRED years ago, when ancient Greece was beginning to carve out the brilliant civilization that was to be her destiny, Thales, the sage of Asia Minor, proposed a search for the answer to the great question which man must always ask, "Of what and how is the world made?" Only by finding a right answer to this question, he contended, is it possible for man to adjust his life and thoughts to his surroundings. It is not surprising that he himself never found a satisfactory solution. For twenty-five centuries we have been continuing his quest, and we are yet far from the complete answer to his problem. Nevertheless progress has been made. The main outlines of our great universe are now spread out before us. The time is ripe for science to report what has been learned and what remains to be found out "concerning the nature of things."

Earth, water, air, and fire were the elements out of which Thales would build his world. His successors however carried the analysis deeper. Democritus and his fellow atomists recognized that things are made of minute particles, which they called atoms, and in terms of these atoms

[1]Delivered at the Rice Institute in March, 1932, by Professor Arthur H. Compton of the University of Chicago.

they accounted for everything from a gorgeous sunset to the love of a man for a maid. This was mostly lost in the dark days of the Middle Ages, and has been rediscovered only during the last few centuries. In that relatively brief era of world history between Copernicus and Einstein we have learned much more reliably and precisely than the Greeks ever knew about man's surroundings.

Let us take a brief preview of the microworld as we now know it. Stars, mountains, people—what are the parts of which they are made, and how are these parts put together? Light, radio rays, X-rays—what are these radiations?

When we take apart this infinitely complex mechanism that we call dirt, or perhaps a diamond, or it may be a flower, we find it made up of a myriad of tiny molecules. Each molecule is itself complex, but is more perfectly formed than the wheels of a Swiss watch. Moreover, these molecules are in continual motion, having continued to run for years without winding and without wear. We find that the molecules which make up matter in its endless variety of forms are themselves built up of a few hundred kinds of atoms. Some of these atoms differ from each other by their chemical properties, others only by their weights.

Our few hundred atoms are themselves made of yet more tiny particles, which become evident through their electric charges. There are found to be two kinds of them, carrying positive and negative electric charges respectively. As nearly as we can tell, all those of any one kind are exactly alike. The positively charged particles, called *protons*, have most of the weight of the atom, while the negatively charged particles, or *electrons*, are the lively little bodies responsible for chemical combinations, electrical conductivity, and the like. By grouping themselves in various ways, these little particles form the various atoms.

Rocks and trees and people do not, however, make up the whole of the universe. How about the sunlight that gives us warmth and makes the trees grow? These rays, though having all the characteristics of waves, are found to consist of minute corpuscles, which are known as *photons*.

Protons, electrons, and photons—these are the elements with which the modern physicist replaces the earth, water, air, and fire of Thales. We sometimes think of standardization as being the distinctive keynote of modern industry. Yet even the most standardized motor car has hundreds of different parts. What then shall we say of the Maker of our universe, who, using but three kinds of pieces, protons, electrons and photons, has made our world with its infinite variety of beauty and life.

MOLECULAR MOTIONS

Let me lead you into this microworld through the lenses of a powerful microscope, as Brown was led into it a century ago while examining some tiny objects suspended in water. "They're alive," he exclaimed. But to make sure he turned his glass on some wax particles suspended in the same way, and found that they also moved. Everything that is small enough moves, he found, unless it is held fast to something solid. These are the *Brownian movements*.

What is the cause of this motion? For many years physicists had explained the pressure of a gas on the walls of a vessel which holds it by supposing that the gas is made of little particles, the "molecules," which dart about in the gas at high speed. They could calculate how fast these molecules must go to account for the observed pressure, and found that if the energy of their motion was proportional to the temperature they could accurately account for the increase of pressure with temperature. The calculation

showed too that the energy of motion of the little molecules should be the same as that of the big molecules, which would mean that the little molecules must move the faster.

Now a careful study of these Brownian movements reveals the fact that the energy of the particles under the microscope is just what the kinetic theory says a molecule should have. And the speed of their motions increases at higher temperatures, just as the molecular theory predicted. At one instant more molecules strike a globule on one side, and the next instant more strike on another. In fact, in a very real sense, one may correctly consider the motions of these little particles as true molecular motions. We must remember, however, that these globule "molecules" which are seen under the microscope consist of perhaps a million or more atoms each, whereas a gaseous molecule usually has only one atom or two or three.

We have here a glimpse of Nature's continual activity. How far from the truth is our idea of "dead matter"! Our little friends, the atoms and molecules, to whom we have just introduced you, are indeed very lively little fellows. If they were not, whence would our own life come?

ATOMS

"According to convention there is a hot and a cold, a sweet and a bitter, and according to convention there is color; in truth there are atoms and a void."

—Democritus.

A few years ago we were camped beside a mountain lake in the foothills of the Himalayas. The warm air from the plains of India was blowing over a nearby mountain range, and was coming down again into the beautiful Vale of Kashmir. Clouds were continually forming as the air, cooled by expansion on coming up the mountain side, became

supersaturated with moisture. But after passing the crest of the range, the air was warmed again by compression as it sank to lower levels, and the clouds evaporated into the thin air.

It was while watching a similar occurrence in his native hills of Scotland that C. T. R. Wilson conceived his beautiful laboratory experiments on clouds. Though he could not bring the mountains into his laboratory, he could expand his moist air in a cylinder of glass and see what was

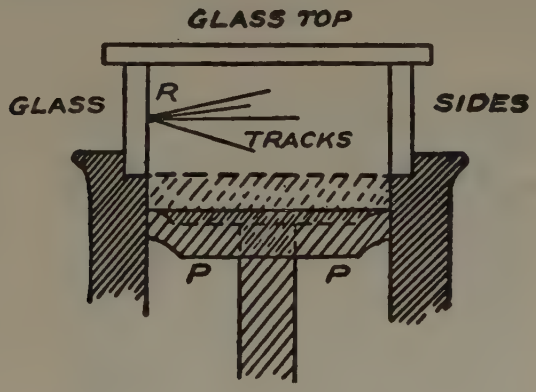


Fig. 1

going on. Figure 1 shows an instrument similar to Wilson's original expansion chamber. Note its glass walls and top and the movable piston at the bottom. Some water is also kept in the chamber, so the air will be moist. After the air is compressed, it is allowed to stand for a moment until it is saturated with moisture, and is then allowed to expand. The resultant cooling causes a cloud to form in the glass chamber just as it did on the mountain top.

Did it ever occur to you that when a cloud forms each little drop of moisture in the cloud must condense on something? Usually it condenses on a speck of dust floating in the air, so that after a rainstorm these dust particles are

carried to the ground and the air is beautifully clear. But when the dust has been removed, on what can the moisture condense? There are always in the air some broken bits of atoms and molecules which we call *ions*. These ions may be produced by rays from radioactive substances in the ground, and other sources. So Mr. Wilson tried the experiment of placing a speck of radium in his expansion chamber, to see what kind of clouds would be formed.

Instead of a diffuse cloud, little white lines of fog appeared, like rays shooting out from the speck of radium. These linear clouds are easily visible, and may be photographed, as in figure 2, which is one of Wilson's earliest cloud pictures. We see the glass walls of the expansion chamber, and the rod on the tip of which the radium is placed. The white lines radiating from this point are clouds of water droplets. Some of the more diffuse lines show their foggy character more clearly.

Why do the clouds form along these radiating lines? It is apparent that the lines mark the paths of some particles shot out from the radium. A photograph taken so as to show the lines more sharply is shown in figure 3.

ATOMS OF HELIUM

What are these particles shot out from the radium? Let us call them alpha particles, in order not to imply anything about what they are, and investigate their properties.

Lord Rutherford once caught a large number of these particles in the effort to learn what they make when there are enough of them to handle. He compressed a highly radioactive gas into a fine glass tube with walls so thin that the alpha particles could pass through. After a few days he noticed gas collecting in the space surrounding the tube. When an electric discharge was passed through the collected

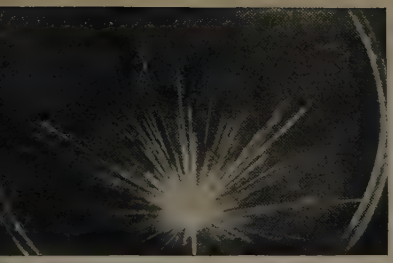


Fig. 2

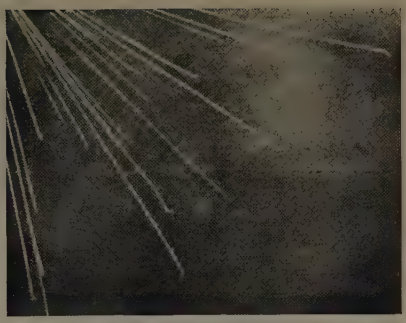


Fig. 3

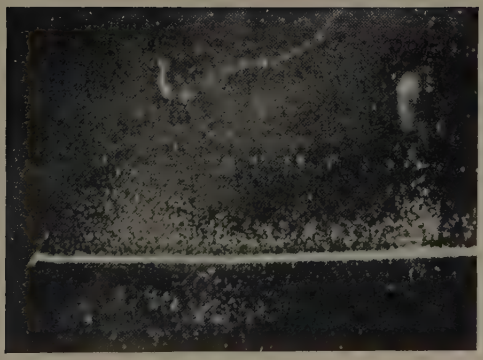


Fig. 4

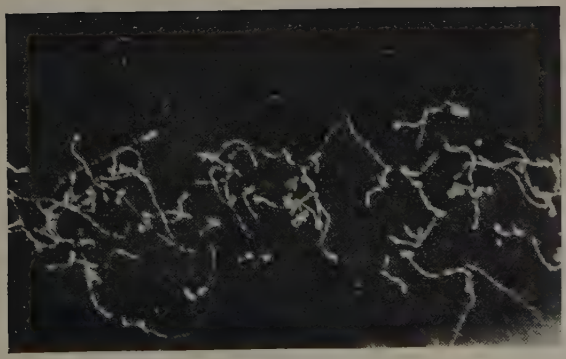


Fig. 5

gas, the light viewed through a spectroscope showed the brilliant spectrum of the gas helium.

Let us recall the romance of helium. Observed many years ago by Lockyer in the spectrum of the sun, it remained unknown on earth for a generation until Raleigh and Ramsay, making a precise measurement of the density of nitrogen in the air, found it different from the nitrogen prepared in the laboratory. Search for the cause of this discrepancy revealed a whole series of new gases—argon, with which our incandescent lamps are filled; neon, with which we advertise our wares in blazing red; helium, with which we now fill our dirigibles; and two others, krypton and xenon, which have become of great value in certain laboratory experiments. Thus was helium found; and in these cloud photographs we see it being formed—the birth of helium atoms. For these alpha particles are none other than atoms of helium gas.

It is possible to count the atoms one by one as they come from a preparation of radium. This might be done with the help of a cloud expansion chamber, counting the tracks as they appear. We should thus have, however, only an intermittent count. A better method is to allow the atoms to enter an electrical counting chamber. They can thus be made to make an audible sound as each particle enters the chamber, or still better, each particle can be made to trip an electrical counter. Rutherford's glass tube, through whose walls the alpha particles came which made up the helium gas, was immensely more radioactive than the radium samples used in such counting experiments. Nevertheless, by comparing the radioactivity of the two samples we can estimate the number of atoms that must have come out of the glass tube. Suppose that we have thus counted all the atoms that have come through the tube and made the gas

observed in Rutherford's spectroscope. How many atoms would we have? In a small thimble filled with helium at atmospheric pressure, the number of atoms is about 1 with nineteen ciphers after it. For helium, each atom is a molecule. In the case of air each molecule of nitrogen or oxygen has two atoms. As Avogadro showed many years ago, however, the number of molecules of air or helium or any other gas required to fill a cubic centimeter at atmospheric pressure is the same for all, and is found to be 1.9×10^{19}. This is known as "Avogadro's number."

Perhaps such a number means but little. Let us then put it this way: Two thousand years ago Julius Cæsar gave a dying gasp, "*Et tu, Brute?*" In the intervening milleniums the molecules of air that he breathed out with that cry have been blown around the world in ocean storms, washed with rains, warmed by the sunshine, and dispersed to the ends of the earth. Of course only a very small fraction of those molecules are now in this room; but at your next breath each of you will inhale several of the molecules of Caesar's last breath.

Molecules and atoms are very tiny things; but there are so many of them that they make up the whole world in which we live.

PARTS OF ATOMS

The story is told of Lord Kelvin, famous Scotch physicist of the last century, that after he had given a lecture on atoms and molecules, one of his students came to him with the question, "Professor, what is your idea of the *structure* of the atom?" "What!" exclaimed Kelvin, "The structure of the atom! Don't you know that the very word 'atom' means the thing that can't be cut? How then can it have a structure?"

"That," remarked the facetious young man, "shows the disadvantage of knowing Greek."

To the ancient atomists the "atom" was the smallest unit into which matter could be divided. The physicist of the present day gives the name of "atom" to a particular type of minute unit of matter which is found to occur in nature. For many years no method was known of dividing these units into smaller components, and they were therefore frequently identified with the "atoms" of the ancients. There are, however, several hundred different kinds of atoms recognized, one for each of the chemical elements, and many of the elements are composed of several different types of atoms which differ from each other in weight. May there not be some more fundamental substance of which these atoms in turn may be built?

THE ELECTRON

In figure 4 is shown a photograph of the trail of helium atom on a larger scale than those shown in figures 2 and 3. Below the trail left by the helium atom we notice a much fainter trail, which is very crooked. This appears to be due to something much smaller than the particle which made the broad bright trail above. If we called the upper one an *alpha* particle, let us call the lower one a *beta* particle, and try to find out what kind of thing it is.

Similar beta particles knocked out of air by the action of X-rays are shown in figure 5. The air in the chamber might be replaced with argon, and beta particles would still appear. If thin sheets of aluminum or gold are placed in the chamber, beta particles are seen to come out of the metal. It is found that they can be ejected from anything. Beta particles are, that is, a common component of all different forms of matter of whatever nature.

The photograph shown in figure 6 was taken with a strong magnetic field applied to the expansion chamber when the cloud was formed. It will be seen that the trails of the beta particles are curved in circles. Such a bending of paths is just what we should expect if the moving particles are electrically charged. For a magnetic field produces a force on a moving electric charge just as it does on a wire carrying an electric current. From the direction of the curvature one can show that the particle is negatively electrified.

Many years ago, Michael Faraday, the centennial of whose great discovery of electromagnetic induction was celebrated last year, showed that when an electric current passes through a liquid, it is carried by electrically charged atoms, or ions, each of which carries one or two or three unit electric charges. Thus the hydrogen ion always carries one unit charge and the oxygen atom two units of charge. So when an electric current is passed through water, it requires twice as many hydrogen atoms to carry the current to the cathode as oxygen atoms to the anode, and hydrogen is produced twice as rapidly as oxygen. Similarly the silver ion carries one unit of charge, whereas the copper ion carries two units. Thus if the same electric current passes through solutions of silver and copper, twice as many silver atoms are liberated as copper atoms. Since we have found how to count the molecules in a unit volume of gas, if we measure the quantity of electricity required to produce a unit volume of hydrogen by electrolysis, we can indeed calculate the charge carried by the hydrogen ion. It is merely $e=Q/2N$, where Q is the total amount of electricity required, N is Avogadro's number of molecules per unit volume, and 2 is the number of hydrogen atoms in a hydrogen molecule. This quantity e is about 0.000,000,000,5 of an electrostatic unit, and is known as the *electronic unit of charge.*

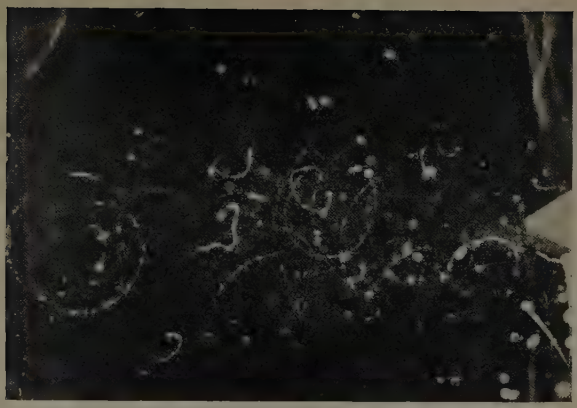


Fig. 6

Fig. 7

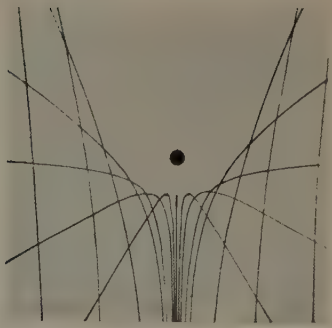


Fig. 8

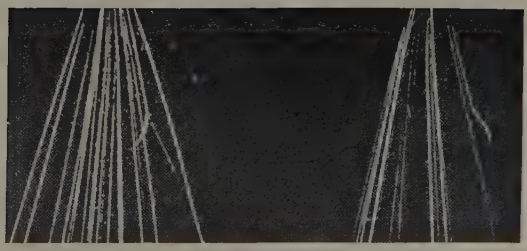


Fig. 9

Millikan has devised an electroscope which is so sensitive that it is capable of measuring even the minute charge carried by a beta particle. In place of the ordinary gold leaf, he caught a tiny drop of oil or mercury between a pair of electrified plates, and threw a beam of ultraviolet light or X-rays on the droplet. The effect of the light was to eject a beta ray. This left the droplet with a positive charge equal to the negative charge carried away by the beta particle, and the electric field due to the plates could be so adjusted that the force on the droplet's charge would just balance its weight. Knowing the weight of the droplet, it was thus possible to determine its charge. It was found that each beta ray which left the droplet carried with it the same negative charge, a charge equal in magnitude to that carried by the hydrogen ion in electrolysis.

The electronic unit of charge is thus of very fundamental significance. It measures not only the charge on the various atomic ions in electrolysis, but also the charge of these beta particles which seem to be sub-atomic in size.

When once we know the charge carried by these particles, we can estimate how heavy they are by measuring how sharply they are bent by a magnetic field. A massive particle will be bent only slightly from its course, whereas a lighter one will be deviated more easily. Magnetic deflection experiments, such as those shown in figure 6, show that a beta particle is much more readily deflected than is a hydrogen ion that carries the same charge. In fact the mass of the beta particle is only 1/1845 that of a hydrogen atom. We were thus right in our guess that these beta particles are things much smaller than the smallest known atom.

Since we find that such particles can be removed from every kind of matter, it follows that they must be one of

the components of which the atoms themselves are built. We shall now give to these beta particles the name of *electrons*.

THE HEART OF THE ATOM

It is not possible to suppose that atoms are constituted solely of electrons, for electrons have negative charges, and atoms are electrically neutral. The atoms must therefore have some positive electric charge as an essential part of their structure. An examination of the trails of alpha particles such as those shown in figure 7 give an important clue to the distribution of the positive charge within the atom. It is really a most remarkable thing that most of the alpha ray tracks are so straight, since a simple calculation shows that in the portion of the tracks shown in this figure each alpha particle has passed through some 20,000 atoms of air. But we have seen that the alpha particle is itself a helium atom with a positive charge. This photograph therefore means that while the helium atom is passing through the oxygen and nitrogen atoms of air, we have repeated 20,000 times the most unusual phenomenon of two bodies occupying the same space at the same time.

Occasionally, however, the alpha particle strikes something so hard and immovable that it must change its course. From the fact that there is only one such collision for thousands of atoms traversed, it is clear the object struck is much smaller than the atom itself. Further, the manner in which the impinging helium atom glances off shows that the object is heavier than an atom of helium. It is not, however, a collision with a stone wall. There is a fork in the track, showing that the particle struck by the alpha particle itself rebounded and left a short track.

From the relative length of the tracks of the deflected alpha particle and the thing which recoiled from the impact,

it seems probable that the object has a mass several times that of a helium atom, or about the mass of an oxygen atom. Collisions of this character indicate that there must be something within the atom hard and impermeable, very much smaller in size than the atom itself, yet possessing practically the whole mass or weight of the atom. This something has been named the atomic *nucleus*.

It has been shown by Lord Rutherford that an atomic nucleus deflects an alpha particle as if the force between them were one of repulsion between two electric charges. On this view the paths of alpha particles passing a nucleus should be as shown diagrammatically in figure 8. All the particles are bent slightly, but only those coming close to the nucleus are bent through a large angle. It is obvious that if the charge on the nucleus is large its effect will extend to a greater distance. That is, the nucleus will act as a larger obstacle and the number of collisions will be greater. So by counting the number of collisions occurring when a group of alpha particles passes through a known number of atoms we can determine the charge on the nucleus. Measurements of this kind have shown that the nucleus of the hydrogen atom has a positive charge equal to one electronic unit, helium that of two, lithium three, and so on down the list of chemical elements to uranium with a charge equal to 92 electrons on its nucleus. This is expressed by saying that the nuclear charge is equal to the *atomic number* of the element.

PROTONS, OF WHICH NUCLEI ARE BUILT

This discovery suggests that the nucleus may be built of units carrying a positive charge equal to the negative charge of the electron. Such a unit we find in the nucleus of the hydrogen atom. It is perhaps surprising that the positive

unit of electric charge should have almost two thousand times the mass that is associated with the negative unit. Rutherford has, however, performed a series of experiments which gives good reason to suppose that our guess is correct. These experiments consist in shooting alpha rays from radium through various substances. It is found that particles having the same charge and mass as the hydrogen nucleus can be knocked out of some of the lighter elements. An event of this kind is shown in figure 9, a remarkable photograph taken by Mr. Blackett. We see here the impact of an alpha particle with the nucleus of a nitrogen atom. There is a thin trail left by the hydrogen nucleus escaping from the nitrogen nucleus.

Similar results are obtained when alpha particles traverse boron, aluminum, phosphorus, and certain other light elements. It would seem that the only reason that hydrogen cannot be liberated from other elements by such methods is that our hammer, the alpha particle, does not strike a sufficiently powerful blow. It would thus appear that the nuclei of the various atoms are indeed built of an aggregate of hydrogen nuclei cemented together with electrons. These hydrogen nuclei we shall now call *protons*, i.e., the fundamental things.

It is the nucleus of an oxygen atom which appears associated with the deflected alpha particle in figure 7. For an oxygen atom of atomic weight 16, this nucleus presumably consists of 16 protons and 8 electrons, leaving a positive charge equal to its atomic number of 8. There are, however, a very few oxygen atoms which have weights of 17 and 18 respectively. Their nuclei must thus consist of 17 or 18 protons and 9 or 10 electrons, respectively, in order that both their masses and their resultant charges shall be correct. For it is the resultant charge of the nucleus, 8 electronic

units, which distinguishes oxygen from any other chemical element.

We have thus found the parts of which atoms are built. There is an electron atmosphere, which is responsible for the ordinary physical and chemical properties of the atom, and inside this atmosphere a tiny massive core, compactly built of protons and a few electrons. It is a remarkable thing that this wonderful world of ours can be composed of only a few hundred kinds of atoms. Now we find that all of these atoms are made of two more fundamental units, protons and electrons. We seem to be approaching the ultimate in simplicity.

II

WHAT IS LIGHT?

AS long ago as the seventeenth century, Newton defended the view that light consists of streams of little particles, shot with tremendous speed from a candle or the sun or any other source. At the dawn of the nineteenth century, however, experiments were performed which gave positive evidence that light consists of waves. Maxwell interpreted them as electromagnetic waves, and in such terms we have ever since been explaining light rays, X-rays, and radio rays. We have measured the length of the waves, their frequency and other characteristics, and have felt that we know them intimately. Very recently, however, a group of electrical effects of light has been discovered for which the idea of light waves suggests no explanation, but whose interpretation is obvious according to a modified form of Newton's old hypothesis of light projectiles.

WHAT WE MEAN BY "LIGHT"

When the physicist speaks of light he thinks not only of those radiations which affect the eye, he refers rather to a wide range of radiations, similar to light in essential nature, but differing in the quality described variously by the terms color, wave-length, or frequency. At one end of this series of radiations are the wireless, or radio rays, with which in recent years we have become so familiar. There is an important point regarding these rays to which I should like to call attention. When one strikes the strings of a mandolin,

they are set into vibration, and produce in the surrounding air the waves which affect our ears and cause the sensation of sound. Investigation shows that the sound waves in air vibrate with the same frequency—the same number of times per second—as do the strings on the mandolin. In precisely the same way, when an electrically charged condenser is discharged an oscillation of the electric charge is set up which gives rise to electric waves, just as the vibrating string produced sound waves. That is, the emitted electric waves have the frequency of the oscillating source.

Though visible light is known to be essentially the same kind of thing as these electric waves, we have long sought in vain for any oscillator which would emit light waves having the same frequency as that of the oscillating source. It was only when Heisenberg introduced a new kind of mechanics, differing radically from the classical ideas of Newton, that we found that the atom vibrates with certain "overtones" whose frequency is that of the light waves which come from it. This is one of the serious difficulties with the wave conception of light, which could only be solved by a fundamental change in our ideas regarding how things work.

Measured in terms of the length of a wave, from one crest to the next, electric waves extend from many miles in length, down through the radio waves of say 300 meters, to the very short waves resulting from tiny sparks, which may be no more than a tenth of a millimeter in length. These rays overlap in wave length the longest heat waves radiated by hot bodies, and may be detected and measured by the same instruments. A familiar source of such heat rays is the reflector type of electric heater, the kind that warms one side of us in a chilly room. The greater part of these heat rays are intermediate in wave-length between

the shortest electric waves and visible light. Such a heater, however, glows a dull red, meaning that its rays extend into the visible regions.

Ordinary light, such as that from the sun, may be spread out into a spectrum by allowing it to pass through a prism. Beyond the red end of the spectrum lie the heat rays. Indeed, if we should place a radiometer just beyond the red end of the spectrum, we should find it strongly affected by the heat rays from the sun. The question arises, are there similar radiations beyond the violet which we are unable to see?

Though the eye is not sensitive to light in this region of the spectrum, a photographic plate placed beyond the violet receives an impression, and the radiation in this region can be made visible by placing in its path some fluorescent substance, such as petroleum oil. These are the ultra-violet rays, of which we have recently heard so much in connection with summer sunshine and the prevention of rickets.

As one goes farther into the ultra-violet, the rays become rapidly absorbed by air, and can be studied only in a vacuum. But at still shorter wave-lengths the rays are again less readily absorbed as we approach the region of X-rays. Just as in the case of the ultra-violet light, these rays do not affect our eyes. They do however affect a photographic plate or produce fluorescence just as does the ultra-violet light. That X-rays are of the same nature as light is shown by the fact that we have been able to refract and reflect them, to polarize and to diffract them. They are indeed light of ten thousand times shorter wavelength.

One of the most important properties of X-rays is their ability to ionize air and make it electrically conducting. This is shown for example by the fact that a charged electro-

scope when exposed to X-rays is promptly discharged. This is due to the breaking up by the X-rays of the oxygen and the nitrogen atoms of the air. Precisely the same thing happens when the atoms in one's body are exposed to X-rays. It is this which makes possible X-ray therapy.

Such ionization can also be produced by the gamma rays from radium. These rays are more penetrating even than X-rays. Whereas X-rays may be half absorbed in an inch of water, it takes a foot of water to absorb half of a beam of gamma rays, corresponding to the much shorter wave-length of the gamma rays.

But the end is not yet. There exists a kind of highly penetrating radiation which is especially prominent at high altitudes, and is supposed to come from some source outside the earth. These *cosmic rays*, as they are called, will penetrate ten or twenty feet of water before they are half absorbed. If these rays are of the same nature as visible light, they must be of yet much shorter wave-length than the gamma rays from radium.

Thus from cosmic rays, with a wave-length of 2×10^{-13} cm. to electric waves 2×10^{6} cm. long there is found to be a continuous spectrum of radiations, of which visible light occupies only a very narrow band. The great breadth of this wave-length range will perhaps be better appreciated if we expand the scale until the wave of a cosmic ray has a length equal to the thickness of a post card. The longest wireless wave would on this scale extend from here to one of the nearer fixed stars.

When the physicist speaks of light, he refers to all the radiations included in this vast range. We believe that they are all the same kind of thing, and that anything which may be said about the nature of the rays in one part of this region is equally true of the rest.

THE WAVE PROPERTIES OF LIGHT

There are many ways in which light acts like a wave in an elastic medium. Such elastic waves move with a speed which is the same for all wave-lengths and all intensities, just as does light. Waves, like light rays, can be reflected and refracted. The polarization of light is a property characteristic of the transverse waves in an elastic solid. It is true that if one examines the constancy of the speed of light in detail, difficulties arise; for it is found that its speed is the same relative to an observer no matter how fast the observer is going. This would not be true if light were a wave in an ordinary elastic medium. Maxwell's identification of light as electromagnetic waves, however, removes this difficulty.

The crucial test for the existence of waves, however, has always been that of diffraction and interference. Imagine that a series of ripples on a pond is passing through the openings of a grid. The crests of the emerging wavelets recombine to form a new wave going straight ahead. But in addition, the wavelet just emerging from one opening may combine with the first wave from the next opening, the second from the next, and so on, forming a new wave-front inclined at a definite angle to the first. The angle between these two waves is determined by the distance between the successive waves—the wave length—and by the distance between successive openings in the grid.

That such a variety of wave formation is not purely imaginary is shown in figure 1, which is a photograph of ripples on the surface of mercury, taken after they have passed through a comb-like grid. Notice how one group of waves combines to form a wave-front going straight ahead. But in addition, on either side of the central beam we find

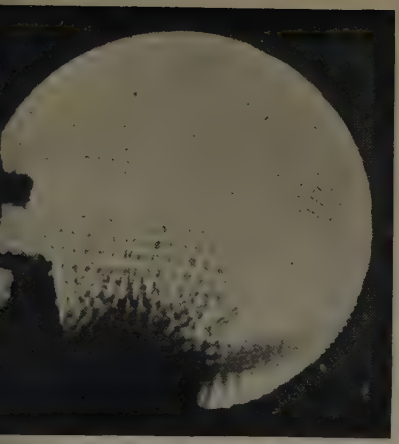


Fig. 1

Fig. 2

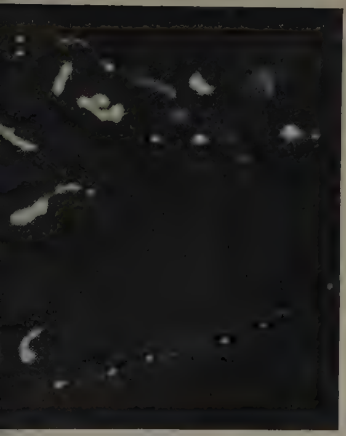


Fig. 3

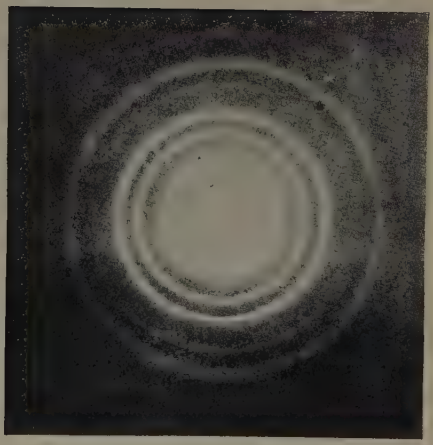


Fig. 4

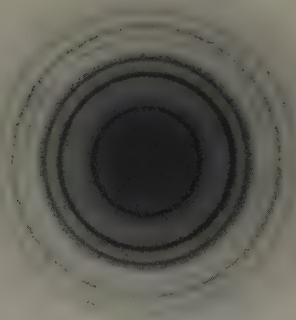


Fig. 5

two beams forming where parts from successive openings in the grid differ by one wave-length. Out at a larger angle we see even a second order of the diffracted beam where there is a difference of two wave-lengths between the ripples coming from adjacent openings.

If we were unable to see the successive waves, but knew the kind of grid through which the ripples had passed, not only could we say that this is the way in which the beam should be split up if it consists of waves, but we could even tell what the wave-length of the ripples must be in order to give these particular angles to the diffracted beam.

During the present lecture we performed the same experiment with a beam of light. A set of some 200 vertical lines was photographed onto a lantern slide, forming a grid through which a beam of light was made to pass. When this grid was placed in front of the lens of the projection lantern, the original line of light projected on the screen was split into three, a bright one in the center—the direct ray—and a diffracted ray on either side. It was just as in the case of the mercury ripples passing through the grid. A grid with about 300 lines to the inch was then placed over the lens, and the separation of the lines was much greater. The outer edges of the diffracted lines were red and their inner edges blue. This means that red light is of the greater wave-length. In fact it would have been possible from this experiment to tell what the wave-length of light is: The distance from the central image to the diffracted image is to the distance from the lantern to the screen as the wave length of the light is to the distance between the lines on the grating. When one carries through the calculation, he finds that the wave-length of the light is about a fifty-thousandth of an inch.

Precisely similar experiments can be done with X-rays. Only in place of the projection lantern we use an X-ray

tube and a pair of slits. The slide with the lines photographed on it is replaced by a polished mirror on which lines are ruled fifty to the millimeter. Instead of the screen we use a photographic plate. A typical resulting photograph is shown in Fig. 2. When the ruled mirror is withdrawn, we have the single vertical line *D*. With the grating in place, we see a bright central reflected image *O* with companions on either side. Thus X-rays can also be diffracted, and must, therefore, like light, consist of waves.

LIGHT CONSISTS OF PARTICLES

For a hundred years no one had seriously questioned the truth of the wave theory. At the close of the last century even the difficulty of supplying a suitable oscillator to give rise to the light waves seemed about to disappear through the discovery of electrons. But in 1900 Planck published the results of a long study of the problem of radiation of heat and light from a hot body. This difficult theoretical study, which has stood the test of time, showed that if a body when heated is to become first red hot, then yellow, and then white, the oscillators in it which are giving out the radiation must not radiate continuously as the electromagnetic theory would demand. They must rather radiate suddenly little portions of energy. The amount of energy in each portion must further, according to Planck, be proportional to the frequency. This is the origin of the celebrated "quantum" theory.

On account of the difficult reasoning involved in Planck's argument, his conclusions carried weight only among those who were especially interested in theoretical physics. Among these was Einstein, who called attention to the fact that Planck's conclusions would fit exactly with the view that the radiation was not emitted in waves at all, but as little

particles, each possessing a portion of energy proportional to the frequency of the oscillator, as Planck had assumed.

EINSTEIN AND THE PHOTOELECTRIC EFFECT

An opportunity to apply this idea was afforded by the photoelectric effect. It is found that when light as from an arc falls upon certain metals, such as zinc or sodium, a current of negative electricity in the form of electrons escapes from the metallic surface. This photoelectric effect is especially prominent with X-rays, for these rays eject electrons from all sorts of substances.

The most remarkable property of these photoelectrons is the speed at which they move. X-rays are produced when cathode electrons bombard a metal target inside the X-ray tube. Let us suppose that the cathode electron strikes the target at a speed of a hundred miles a second—quite a normal speed for these little particles. The resulting X-ray, after passing through the walls of the tube and perhaps a block of wood, may eject a photoelectron from a metal plate placed on the far side. The speed of this photoelectron is then found to be almost as great as that of the original cathode electron.

The surprising nature of this phenomenon can perhaps be emphasized by an allegory. When I was a young lad, my father used to take our family to a lake in northern Michigan. My older brother, with several of the older boys, built a diving pier beyond the point half a mile away from camp, where the water was deep. Fearing lest something would happen, mother would not allow us younger boys to swim in this deep water. So we built a diving pier of our own in the shallower water in front of camp. It so happened, one hot, calm, July day, that my brother dove from his diving board into the deep water. The ripples from

the resulting splash must have spread out over the lake. By the time they had gone around the point to where I was swimming, they were of course much too small to notice. You can imagine my surprise, therefore, when these insignificant ripples, striking me as I was swimming under our diving pier, suddenly lifted me from the water and set me on the diving board!

Of course this story is impossible. Yet if it is impossible for a water ripple to do such a thing, it is just as impossible for an ether ripple, sent out when an electron dives into the target of an X-ray tube, to jerk an electron out of a second piece of metal with a speed equal to that of the first electron.

It was considerations of this kind that showed to Einstein the futility of trying to account for the photoelectric effect on the basis of waves. He saw, however, that this effect might be explained if light and X-rays consist of particles. These particles are now commonly called *photons* or *light quanta*. The picture of the X-ray experiment on this view would be that when the electron strikes the target of an X-ray tube, its energy of motion is transformed into a photon, that is, a particle of X-ray, which goes with the speed of light to the second piece of metal. Here the photon gives up its energy to one of the electrons of which the metal is composed, and throws it out with an energy of motion equal to that of the first electron.

In this way Einstein was able to account, in a very satisfactory way, for the phenomenon of the ejection of electrons by light and X-rays.

PECULIAR X-RAY ECHOES

Even more direct evidence that light consists of particles has come from a study of X-ray echoes. If you hold a piece of paper in the light of a lamp, the paper scatters

light from the lamp into your eyes. In the same way, if the lamp were an X-ray tube, the paper would scatter X-rays into your eyes. If light and X-rays are waves, scattered X-rays are like an echo. When one whistles in front of a wall, the echo comes back with the same pitch as the original sound. This must be so; for each wave of the sound is reflected from the wall, as many waves return as strike, and the frequency or pitch of the echoed wave is the same as that of the original wave. In the case of scattered X-rays, the echo should similarly be thrown back by the electrons in the scattering material, and should likewise have the same pitch or frequency as the incident rays.

A few years ago we measured the pitch of some X-ray echoes, using an X-ray spectrometer. We found that though a part of the scattered rays is of the original wave-length, a greater part is of increased wave-length. This would correspond to a lower pitch for the echo than for the original sound.

As we have seen, this change in wave-length is contrary to the predictions of the wave theory. If we take Einstein's idea of X-ray particles, however, we find a simple explanation of the effect. On this view we may suppose that each photon of the scattered X-rays is deflected by a single electron. Picture to yourselves a golf ball bouncing from a football. A part of the golf ball's energy is spent in setting the football in motion. Thus the golf ball bounces off having less energy than when it struck. In the same way, the electron from which the X-ray photon bounces will recoil, taking part of the photon's energy, and the deflected photon will have less energy than before it struck the electron. This reduction in energy of the X-ray photon corresponds, according to Planck's original quantum theory, to a decrease in frequency of the scattered X-rays, just as the

experiments show. In fact, the theory is so definite that it is possible to calculate just how great a change in frequency should occur, and the calculation is found to correspond accurately with the experiments.

RECOILING ELECTRONS

If this explanation is the correct one, it should however be possible to find the electrons which recoil from the impact of the X-ray particles. Before this theory of the origin of scattered X-rays was suggested, no such recoiling electrons had ever been noticed. Within a few months after its proposal, however, C. T. R. Wilson succeeded in photographing the trails left when electrons in air recoil from the X-rays which they scatter. Figure 3 shows one of his typical photographs. The X-rays are going from left to right. At top and bottom you notice the long trails left by two photoelectrons which have taken up the whole energy of a photon. In between are a number of shorter trails, all with their tails toward the X-ray tube. These are the electrons which have been struck by flying X-ray photons. Some have been struck squarely, and are knocked straight ahead. Others have received only a glancing blow, and have recoiled at an angle. Thus we have observed not only the loss in energy of the deflected photons, as shown by the lowering in pitch of the X-ray echo, but we have found also the recoiling electrons from which the photons have bounced.

In order, however, to satisfy ourselves by a crucial test whether X-rays act like particles, an experiment was devised which would enable one to follow both the photon as it is deflected by an electron and the motion of the recoiling electron. So feeble a beam of X-rays was used that on the average only one or two recoil electrons appear at a

time. Let us suppose that the first electron struck by the X-ray particle recoils downward. This must mean that the X-ray particle has been deflected upward. If this X-ray should strike another electron before it leaves the chamber, this event must occur in a definite upward direction. It cannot occur on the same side as the recoil electron. If, however, the X-ray is a wave, spreading in all directions, there is no more reason why the second electron associated with the scattered ray should appear on one side than on the other. A series of photographs which show the relation between the direction of recoil of the first electron and the location of the second electron struck by the scattered X-ray thus affords a crucial test between the conception of X-rays as spreading waves and as particles.

From a large number of photographs taken in this manner it has become evident that *an X-ray is scattered in a definite direction like a particle.*

But if X-rays, so also the rest of the family of electromagnetic radiations, for they are all the same kind of thing. It would thus seem that by these experiments Einstein's notion of light as made of particles is established.

THE PARADOX OF WAVES AND PARTICLES

We thus seem to have satisfactory proof from our diffraction experiments that light consists of waves. The photoelectric and scattering experiments afford equally satisfactory evidence that light consists of particles. How can these apparently conflicting ideas be reconciled?

ELECTRON WAVES

Before attempting to answer this question, let me call to your attention the fact that this dilemma applies not only to radiation but also in other fundamental fields of physics.

When the evidence was growing strong that radiation, which we had always thought of as waves, had also the properties of particles, Prince L. de Broglie of Paris asked himself, may it not then be possible that electrons, which we know as particles, have the properties of waves? An extension of Planck and Einstein's quantum theory enabled him to calculate what the wave-length corresponding to a moving electron should be. In photographs such as figure 3 we have ocular evidence that electrons are very real particles indeed. Nevertheless, de Broglie's seemingly absurd suggestion was promptly subjected to experimental test by Davisson and Germer at New York, and later by Thomson at Aberdeen, and others.

You will recall that our crucial evidence for the wave character of light was the fact that light could be diffracted by a grating of lines ruled on glass. X-rays were diffracted in the same way; but before this had been shown possible, it was found that X-rays could be diffracted by the regularly arranged atoms in a crystal. The layers of atoms took the place of the lines ruled on glass. X-rays may be passed through a pair of diaphragms and a mass of powdered crystals which diffract them, and produce an image on a photographic plate. In figure 4 is shown a photograph thus obtained, when X-rays pass through a sheet of aluminum which consists of minute aluminum crystals. The diffraction haloes around the central image form one of the best proofs we have of the wave nature of X-rays.

G. P. Thomson has performed a precisely similar experiment, except that the X-ray beam was replaced with a stream of electrons, and the sheet of aluminum with a gold leaf. A photograph thus obtained is shown in figure 5. Here again are the central image and several haloes produced now by diffracted electrons. If figure 4 demonstrated

the wave character of X-rays, does not figure 5 prove equally definitely the wave character of electrons?

In a similar way recent experiments by Dempster have shown that protons can be diffracted by crystals. Johnson has diffracted neutral hydrogen atoms. Stern has done the same with helium atoms and hydrogen molecules. It would seem that all of these "particles" have wave characteristics if an appropriate experiment is performed to detect them.

We are thus faced with the fact that the fundamental things in nature, matter and radiation, present to us a dual aspect. In certain ways they act like particles, in others like waves. The experiments tell us that we must seize both horns of the dilemma.

A SUGGESTED SOLUTION

There has gradually developed a solution of this puzzle, which, though at first rather difficult to grasp, seems to be free from logical contradictions, and capable of describing the phenomena which our experiments reveal. The point of departure is the mathematical proof that the dynamics of a particle may be expressed in terms of the propagation of a group of waves. That is, the particle may be replaced by a wave train—the two, as far as their motion is concerned, may be made mathematically equivalent. The motion of a particle such as an electron or a photon in a straight line is represented by a plane wave. The wavelength is determined by the momentum of the particle, and the length of the train by the precision with which the momentum is known. In the case of the photon, this wave may be taken as the ordinary electromagnetic wave. The wave corresponding to the moving electron is usually called by the name of its inventor, a de Broglie wave.

It is not usually possible to describe the motion of either a beam of light or a beam of electrons without introducing both the concepts of waves and particles. There are certain localized regions in which at a certain moment energy exists, and this may be taken as a definition of what we mean by a particle. But in predicting where these localized positions are to be at a later instant, a consideration of the propagation of the corresponding waves is usually our most satisfactory mode of attack. According to this theory, electromagnetic waves and the de Broglie waves are both waves of probability. That is, they tell where the particles probably are.

Consider as an example the diffraction pattern of a beam of light or of electrons, reflected from a ruled grating, and falling on a photographic plate. In the intense portion of the diffraction pattern there is a high probability that a grain of the photographic plate will be affected. In corpuscular language, there is a high probability that a photon or electron, as the case may be, will strike this portion of the plate. Where the diffraction pattern is of zero intensity, the probability of the particle striking is zero, and the plate is unaffected. Thus there is a high probability that a photon will be present where the intensity of the electromagnetic wave is great, and a lesser probability where this intensity is smaller.

It is a corollary that the energy of the radiation lies in the photons and not in the waves. For we mean by energy the ability to do work, and we find that when radiation does anything it acts in particles.

Thus we find from our experiments on diffraction and interference that light consists of waves. The photoelectric effect and the scattering of X-rays give equally convincing reasons for believing that light consists of particles. For

centuries it has been supposed that the two conceptions are contradictory. Goaded on, however, by the obstinate experiments, we seem to have found a way out. Whenever the light does anything it works as particles; but in predicting where the particles are to appear we continue to think of the light propagated as waves.

III

THE RIDDLE OF COSMIC RAYS

"UNIQUE in modern physics for the minuteness of the phenomena, the delicacy of the observations, the adventurous excursions of the observers, the subtlety of the analysis, and the grandeur of the inferences," thus K. K. Darrow describes the study of cosmic rays. It is a subject which offers unaccustomed difficulties in public presentation. The first question that is usually asked is, "What are cosmic rays?" To this we must answer, "We do not know." The next question is, "Where do they come from?" Again the answer is, "We do not know." Their existence is however shown by their effects, and their properties lead us to suspect that they may be bringing us messages of unusual interest.

The presence of cosmic rays is recognized through the electrical conductivity that they impart to air and other gases. This is a property shared in common with X-rays, and the alpha, beta, and gamma rays of radium. That the air is conducting has been recognized ever since the time of Coulomb, a century and a half ago. This he proved by the fact that an electroscope slowly loses its charge even though no current passes through the insulation that supports the gold leaves. Electroscopes precisely similar in principle to that of Coulomb, though of course with various refinements, still form one of the chief means of studying the electrical conductivity of the air.

After the discovery of X-rays and radioactivity, it was supposed that this conductivity of the air was due to rays

from radioactive matter distributed in the ground and the atmosphere. Rutherford and Cooke at McGill and McClennan at Toronto tested this hypothesis by surrounding their electroscopes with lead shields. They found as they had anticipated that these shields diminished the conductivity of the air in the electroscopes, due to the absorption of the gamma rays from external radioactive materials. However, after several centimeters of lead had been placed around an electroscope, additional lead produced but little effect. The remaining conductivity was accordingly ascribed to radioactive impurities in the walls of the electroscope itself. Thus it was supposed that the air's conductivity was completely explained.

If this interpretation was the correct one, an electroscope carried above the surface of the earth should register lower conductivity than at the surface. For the radioactive materials which make it conducting are either in the earth itself or in the atmosphere. The rays from radioactive materials in the earth should be absorbed by the lower layers of air and be weaker when they reach an elevated electroscope. Moreover the only radioactive substances in the atmosphere are gases of very high molecular weight, and these should be relatively more abundant at low altitudes than at higher levels. A series of experiments was accordingly carried out to see how the conductivity of the air in an electroscope would vary as it was carried above the ground.

In the first of these experiments the electroscope was carried from the bottom to the top of the Eiffel Tower. It was found, indeed, that its rate of discharge was slower at the higher levels; but the difference was considerably less than had been expected. This unlooked for result stimulated Gockel, Hess, and Kohlhörster to carry on a

series of balloon measurements to study the variation of these rays at greater altitudes. It was found that instead of a continuous decrease in intensity, at about a thousand meters above the earth the conductivity of the air reached a minimum, and began to rise at higher altitudes. At 5350 meters Hess found the conductivity several times that at the earth's surface. At 9300 meters (almost six miles) Kohlhörster reports it "many times" as great. Last year's balloon flight by that intrepid Belgian physicist, Piccard, showed that at 16,000 meters, or ten miles, the conductivity of the air increases to a hundred times its value at the earth's surface. Obviously such effects cannot be due either to radioactivity of the earth itself or to a dense radioactive gas in the earth's atmosphere.

In order to account for this rapid increase in conductivity with altitude, Hess suggested in 1912 that it was probably due to highly penetrating radiation entering the earth from outside its atmosphere. Finding no appreciable variation in its intensity at different times of day, he believed that this radiation was probably coming from some remote portion of the cosmos.

One of the most convincing experiments in showing that the conductivity observed in the air is due in large measure to rays from above was performed a few years ago by Millikan and Cameron. They carried an electroscope to a high mountain lake, and observed the conductivity of its air when submerged at various depths below the surface of the water. At the lower depths the conductivity decreased due to the absorption of the ionizing rays by the water. Then they carried the instrument to a lake at a lower altitude, and repeated the experiment. When the electroscope was lowered to a depth in the upper lake such that the total pressure of the atmosphere and the water was the same as

in the lower lake, the conductivity in the two locations was found to be the same. That is, as they went down the mountain, the conductivity of the air in the electroscope decreased just as it should if the rays producing the conductivity were being absorbed by the layer of atmosphere through which they were passing. This indicated that the conductivity was due to rays coming from some point higher than the higher lake.

Though Hess's experiments would seem to have been enough, it was these experiments perhaps more than any others which convinced physicists of the real existence of a radiation of a highly penetrating kind coming from high altitudes. American newspapers began to speak of them as the "Millikan rays," not knowing that for more than a decade the same rays had been spoken of in Europe as the "Hess rays." Millikan himself seems first to have been convinced of their reality by these experiments, and gave to them their most commonly used name of *Cosmic Rays*.

NATURE OF THE COSMIC RAYS

We have remarked that the conductivity that cosmic rays give to air is a property shared in common with X-rays and the rays from radioactive substances. Of the latter rays, the alpha rays may be completely absorbed by a piece of paper, and the beta rays will penetrate only a few millimeters of solid matter. The gamma rays, however, can be easily detected through a brick wall. It is thus natural to associate the cosmic rays with the gamma rays from radium. Like them they produce ions in gases, thus making the gases conducting, and like them they are very penetrating. In fact the cosmic rays will penetrate from five to a hundred times as much matter as will the most penetrating form of gamma rays. A closer inspection shows that this similarity

to gamma rays holds also for other characteristics. For both types of rays the ionization produced in gases is the same function of the pressure and temperature, and depends in the same way upon the material of the walls. Neither kind of ray is noticeably affected by a magnetic field, though the beta rays and the alpha rays from radium are so affected. Since gamma rays are known to consist of photons, i.e., uncharged particles of the same kind as light, associated with electromagnetic waves, the hypothesis usually made is that cosmic rays are likewise photons.

Whatever the nature of the original cosmic rays, we have direct evidence that the ionization which they produce in air is due to ionizing particles, which appear very much like high speed electrons. This is perhaps best shown by experiments with counting tubes. A counting tube is a device which responds with a sudden rush of electricity every time an ion or a group of ions is produced within it. These counting tubes serve as sensitive means for detecting the presence of cosmic rays. Two or three such tubes are used at a time, the tubes being connected to a single amplifier in such a manner that if an impulse occurs in only one tube nothing happens, but if impulses occur in all of them simultaneously a record is automatically made of the event. The cause of such a coincidence is that an ionizing particle shoots at high speed through all the tubes. If two tubes are placed side by side, comparatively few such "coincidences" occur. If however one is placed over the other, the coincidences are much more frequent. This indicates that the ionizing particles are moving in a vertical direction. Dr. Mott-Smith and Mr. Locher at the Rice Institute have confirmed this conclusion by placing a pair of these counting tubes over a cloud expansion chamber of the type described in the first lecture of this series for making visible the paths of atoms

and electrons. They found that when such a vertical coincidence is observed, a vertical track appears at the same time in the expansion chamber.

That these coincidences and vertical tracks are due to the cosmic rays is shown by the manner in which they are absorbed. A block of lead placed beneath the counters has no effect on the coincidences, whereas if the block is placed above, it reduces the number of coincidences in about the same ratio that the same thickness of lead surrounding an electroscope reduces the conductivity due to cosmic rays. We thus conclude that the ions produced by cosmic rays result from the passage of high speed ionizing particles.

Photons do not, to the best of our knowledge, produce ions in this manner. Being electrically neutral, they traverse large thicknesses of matter without losing any of their energy; but when at last they do strike something, they give up a large part or all of the energy they possess. In doing so they give rise to the photoelectrons and recoil electrons described in the second lecture of this series. These electrons in turn ionize the air as they pass through it, until all their energy is spent. The observation that cosmic ray ionization is due to ionizing particles does not thus exclude the view that the original cosmic rays were photons. For though photons would not themselves produce the effects we observe, they would give rise to recoil electrons which might in turn produce the coincidences and the trails in the expansion chamber.

At this point, however, we must describe a beautiful experiment of Bothe and Kohlhörster which has been supposed to rule out the possibility that cosmic rays are photons. They placed two counting tubes, such as we have described, one above the other, and placed a heavy absorbing block of gold *between* the tubes. They found that the coincidences

were reduced in practically the same ratio as the cosmic ray ionization due to the same thickness of gold. If the cosmic rays were photons, these coincidences must have been due to passage of the recoil electrons through both chambers. With the gold in between, these electrons should however have been almost all absorbed before they reached the second chamber. For experiments with gamma rays show that the recoil electrons which they produce are much more absorbable than the original gamma rays. The experiment showed that the particles which give rise to the coincidences are absorbed at about the same rate as the cosmic rays themselves. Thus they inferred that the particles must be the cosmic rays.

Though in the case of radium the gamma ray or photon is much more penetrating (about a hundred times) than the recoil electron or beta ray which it ejects from matter, the theories of the absorption of these rays indicate that for photons of sufficiently high energy the penetrating power of the recoil electron should approach close to that of the parent photon. It would not in fact require a great change in our theories to make the two practically the same for the very high energies that the cosmic rays particles are known to have. For this reason the experiment of Bothe and Kohlhörster can hardly be considered a crucial one.

Moreover there are serious difficulties involved in interpreting the ionizing particles which give rise to the coincidences as the original cosmic rays which enter the earth's atmosphere from outside. The tracks which appear in Mott-Smith's and Locher's photographs are like those that would be produced by an electrically charged electron or proton moving with a speed nearly equal to that of light. Electrically neutral particles such as photons or neutrons should not show any trail at all. If the particles are charged,

they should be deflected by a magnetic field. Though attempts to deflect these particles by means of powerful electromagnets have been made in vain by Mott-Smith and by Rossi, Curtis has reported a slight effect, and reports from Millikan's laboratory say that they have succeeded in deviating these particles with powerful magnets. Thus it would seem that the ionizing particles are indeed electrically charged, but are moving with very high energy.

If, however, the original cosmic rays are electrically charged, they should be deflected by the earth's magnetic field as they approach the earth. Because of the large size of the earth, the motion of the electron is affected while it is yet a great way off. In fact it has been shown by Epstein that unless the electron has greater energy than that acquired by a fall through ten billion volts it could only strike the earth near the north or south magnetic poles. Various experimenters have tried to find variations in the cosmic ray intensity at different magnetic latitudes on the surface of the earth. Though there is some variation between the results of different investigators, there is certainly no considerable dependence of cosmic ray intensity on latitude. This can only be reconciled with the view that the cosmic rays consist of electrons if we suppose that the energy of these electrons is very great indeed, or that they come not from remote distances, but from the earth's upper atmosphere. In fact the energy required to bring these electrons from interstellar space undeviated by the earth's magnetic field is so great that there is no known source from which it could be derived.

Thus difficulties confront us whether we suppose that the cosmic rays are electrified particles or neutral particles. It is to be hoped that some of the experiments now under way will enable us to choose between the two alternatives.

In the meantime we can say with definiteness that whether photon, electron, or some other kind of particle, the energy associated with each cosmic ray is very much greater than that associated with any of the other rays with which we are familiar. This great energy makes us feel that the solution of the question of their origin and nature is of fundamental importance.

SOURCE OF THE COSMIC RAYS

In searching for a possible origin of the cosmic rays, we note that the fact that the rays are of almost equal intensity day and night, and at different geographic latitudes, means that the origin must be approximately equally distributed in all directions. The earth's atmosphere satisfies this condition. Though the mountain and balloon observations indicate that the source of the rays must be higher than ten miles, there are many interesting atmospheric phenomena which occur at still higher altitudes. There is for example the earth's aurora, which is for the most part more than 100 miles high, and the Kennelly-Heaviside layer of conducting air, so important in radio transmission, which is about 50 miles high. We are at a loss however to suggest any way in which radiations with the characteristics of cosmic rays could be produced in the upper atmosphere. C. T. R. Wilson's suggestion that they might be associated with the electrical discharges occurring in thunderstorms seems to be ruled out by the fact that no significant variation in the intensity of the cosmic rays is found to occur during a storm. We accordingly look farther, keeping in mind however that we have not shown that the atmosphere may not be the origin of these mysterious rays.

It is clear that neither the sun itself, nor any other single heavenly body can be the source of the cosmic rays, for if

it were, we should have a large diurnal variation which is in fact not found. Kohlhörster at one time reported a considerable maximum of intensity when the Milky Way is overhead, and thought this meant that the rays are coming from the stars. But later experiments failed to confirm his experimental results, so that this possibility seems to be ruled out. Millikan has suggested that the rays are produced in the cold regions of interstellar space, which would agree well with their uniform intensity at different times of day.

Within the last year, however, Hess and Lindholm and the writer, all working independently, have found a small but apparently definite maximum in the cosmic ray intensity at noon, and a corresponding minimum at midnight. If this is real, it means that the sun has something to do with the cosmic radiation. Thus, for example, if the cosmic rays originate in the portion of space within an effective radius equal approximately to the radius of the solar system, a slight variation of this kind could be explained.

If the cosmic rays come from interstellar space as Millikan has assumed, they would seem to be of very considerable significance in cosmic evolution. For Millikan has pointed out that the energy in the cosmic rays is comparable with that of starlight. Thus if we receive a fair sample of the rays in interstellar space, the cosmic ray energy in the world must be of the same order of magnitude as the heat and light energy radiated by the stars. MacMillan has indeed suggested that this cosmic ray energy forms ultimately the source from which the stars draw their energy during at least a part of their life cycle. It would thus appear that our understanding of the development of the universe is far from complete unless we include the cosmic rays along with the heat energy.

If, on the other hand, the cosmic rays are a local disturbance started in our own atmosphere, we should thus be attaching to them far too great a world significance. Several experiments are now under way with the objective of learning whether such a local explanation is possible.

HOW ARE THE COSMIC RAYS PRODUCED?

From whatever place the cosmic rays come, the process which produces them must be very energetic. If the rays are photons, in order to account for their penetrating power it would seem that each ray must have from thirty million to a billion electron volts of energy (an electron volt is the energy acquired by an electron in falling through a potential difference of one volt). Such energies may be accounted for as due to either the formation of heavier elements out of hydrogen, or to the annihilation of hydrogen atoms themselves. If on the other hand the rays are electrons or protons, the observed penetrations would seem to imply energies as great as several billion electron volts.

It is possible to calculate the energy involved in an atomic change by use of Einstein's equation relating energy and mass. According to this relation, the energy associated with a mass m is mc^2, where m the mass is expressed in grams, c is the velocity of light expressed in centimeters per second, and the energy is in ergs. Using this relation we find that when the mass of a particle changes by as much as the mass of a hydrogen atom, the energy released is almost exactly a billion electron volts.

Millikan has pointed out that if the cosmic rays are photons, the penetrating power of the softest known component of the rays corresponds to about thirty million electron volts, which is just what should be released if four hydrogen atoms of atomic weight 1.008 suddenly coalesce

to form a helium atom of weight 4.00. Likewise he ascribes some of the intermediate components to the formation of oxygen, silicon, iron, etc. Thus he pictures the cosmic rays as coming from interstellar space, where heavy atoms are being built from lighter components.

Jeans, on the other hand, also using the photon hypothesis, notes that the most penetrating cosmic rays must have an energy approximately equal to that released when a proton or a hydrogen atom is annihilated. This is a process which he has favored to account for the energy of the stars. The softer components he interprets as secondary rays resulting from these more penetrating primary rays. Jeans thus sees in the cosmic rays a stage of the gradual dissipation of the initial energy of the universe.

If we take the alternative view that the cosmic rays are electrons, or some other form of electrified particle, we are at a loss to know how to account for the immense amount of energy (speaking on an atomic scale) that each ray must possess. We might suppose that some heavy atomic nucleus suddenly disintegrates, and spends all of its energy in ejecting one high energy particle. But such a heavy nucleus we believe to be made of many protons and electrons, and it would be surprising if all of these particles should agree upon a suicide pact and annihilate themselves in one act. No other process has however as yet suggested itself whereby such high speed electrified particles might be produced.

We are thus unable to offer a satisfactory account of the origin of the cosmic rays. It seems likely, however, that they result from some change occurring in an atomic nucleus, for this is our only known source of energy which can come in sufficiently large units. Thus we hope that their further study may enable us to gain new information regarding the

structure of that nucleus. In any case, we have here evidence of some kind of process which is liberating immense amounts of energy. We cannot rest until we know where that energy comes from, and find out if possible whether it can be made available to man.

Arthur H. Compton.

CONTENTS

Vol. XIX. January, 1932 No. 1

PAGE

ELEMENTS OF THE THEORY OF RESONANCE ILLUSTRATED BY THE MOTION OF A PENDULUM—A course of lectures given at the Rice Institute, April 22, 23, and 24, 1931, by ERNEST WILLIAM BROWN, D.Sc., F.R.S., Josiah W. Gibbs Professor of Mathematics in Yale University.

Introduction 1

I. The Equation $\frac{d^2x}{dt^2}+\kappa^2 \sin x=0$ 11

Method of Approximation 18

The Generalized Equation $\frac{d^2x}{dt^2}+f'(x)=0$ 20

II. Solutions of $\frac{d^2x}{dt^2}+f'(x)=m\phi'(x, t)$ 21

III. Application to the Pendulum with an Oscillating Support 36

IV. Applications to the Mutual Influence of Two Pendulums 43

V. Solutions of $\frac{d^2x}{dt^2}+\kappa^2x=m\phi'(x, t)$ 53

VI. The Effect of Friction 56

Vol. XIX. April, 1932 No. 2

UNIVERSITY EXTENSION LECTURES ON GOETHE IN OBSERVANCE OF THE ONE HUNDREDTH ANNIVERSARY OF HIS DEATH, ON MARCH 22nd, 1832—A Series of Lay Lectures Delivered at the Rice Institute in the Winter of 1932.

I. Goethe and Philosophy, by RADOSLAV ANDREA TSANOFF, Ph.D. (Cornell), Professor of Philosophy 61

II. Goethe and Literary Criticism, by ALAN DUGALD McKILLOP, Ph.D. (Harvard), Professor of English 84

III. The Correspondence and Conversations of Goethe, by MARCEL MORAUD, Agrégé de l'Université de France, Professor of French 109

IV. Goethe and Science, by ASA CRAWFORD CHANDLER, Ph.D. (California), Professor of Biology 129

V. Goethe and Shakespeare, by STOCKTON AXSON, Litt.D. (Pittsburgh), L.H.D. (Wesleyan), LL.D. (Knox), Professor of English Literature. . . 148

VI. Goethe, Sage and Poet, by HEINRICH MEYER, Ph.D. (Freiburg), Instructor in German 169

Vol. XIX. July, 1932 No. 3

A MATHEMATICAL THEORY OF AESTHETICS AND ITS APPLICATION TO POETRY AND MUSIC—A course of lectures, under the auspices of the Rice Institute Lectureship in Music, delivered in the Physics Amphitheatre, January 4, 5, and 6, 1932, by GEORGE DAVID BIRKHOFF, Ph.D., Sc.D., Professor of Mathematics at Harvard University.

I. Relation of the Theory to Earlier Aesthetic Theories 189

II. The Musical Quality in Poetry 216

III. The Diatonic Chords 242

IV. Harmony 294

V. Melody. 320

Vol. XIX. October, 1932 No. 4

THE FIRE-BRINGER—A sermon preached to the seventeenth graduating class of the Rice Institute, by the REVEREND GEORGE ARTHUR BUTTRICK, D.D., Minister of the Madison Avenue Presbyterian Church of New York 343

SOME ANALOGIES FROM HISTORY—An address delivered at the seventeenth commencement convocation of the Rice Institute by ROSCOE POUND, Ph.D., LL.M., LL.D., Dean of the Law School of Harvard University. 352

The Rice Institute Pamphlet is published at Houston, Texas, and is issued in January, April, July, and October. It was entered as second-class matter, April 15, 1915, at the post-office at Houston, Texas, under the Act of August 24, 1912